Bochel, A. J., Williams, C., McCoy, A. J., Hoppe, H-J., Winter, A. J., Nicholls, R. D., Berger, I., & Crump, M. P. (2020). Structure of the Human Cation-Independent Mannose 6-Phosphate/IGF2 Receptor Domains 7-11 Uncovers the Mannose 6-Phosphate Binding Site of Domain 9. *Structure*. https://doi.org/10.1016/j.str.2020.08.002

University of Bristol - Explore Bristol Research
General rights

Structure of the human cation-independent mannose 6-phosphate/ IGF2 receptor domains 7-11 uncovers the high affinity mannose 6-phosphate binding site of domain 9.

Alice J. Bochel[1,†], Christopher Williams[1,6,†] Airlie J. McCoy[2,], Hans-Jürgen Hoppe[3,7], Ashley J. Winter[1], Ryan D. Nicholls[1], Karl Harlos[4], E. Yvonne Jones[4], Imre Berger[5], A. Bassim Hassan[3,]* and Matthew P. Crump[1,6,]*

1. School of Chemistry, Cantock's Close, University of Bristol, Bristol BS8 1TS, UK.
2. Cambridge Institute for Medical Research, Department of Haematology, University of Cambridge, The Keith Peters Building, Hills Road, Cambridge CB2 0XY, England.
3. Tumour Growth Control Group, Oxford Molecular Pathology Institute, Sir William Dunn School of Pathology, University of Oxford, Oxford, OX1 3RE, UK.
4. Cancer Research UK Receptor Structure Research Group, Division of Structural Biology, Wellcome Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN UK.
5. School of Biochemistry, University of Bristol, University Walk, Bristol, BS8 1TD, UK.
6. BrisSynBio, Life Sciences Building, Tyndall Avenue, Bristol, BS8 1TQ, UK
7. dAInomics Ltd, ᶜ/ₒ Diagnox, 77 Heyford Park, Upper Heyford, Oxfordshire, OX25 5HD, UK

[†]Joint first authors.

*To whom correspondence should be addressed: matt.crump@bristol.ac.uk Phone +44 117 331 7163 Fax +44 117 925 1295 or bass.hassan@path.ox.ac.uk Phone +44 1865 275044 Fax +44 1865 222431.

Summary:

The cation-independent mannose 6-phosphate (M6P)/Insulin-like growth factor-2 receptor (CI-MPR/ IGF2R) is a ~300 kDa transmembrane protein responsible for trafficking M6P-tagged lysosomal hydrolases and the internalisation of Insulin-like growth factor 2 (IGF2). The extracellular region of the CI-MPR is composed of 15 homologous domains but how it interacts with extracellular ligands at neutral pH is poorly understood. We have focused structural work on key domains of human CI-MPR and report the first structures of human domains (D) 7, 8, 9 and 10 within two multi-domain constructs, D9-10 and D7-11. Together these structures provide the first high-resolution description of the high-affinity M6P binding domain 9. Domain 9 stabilizes a well-defined hub formed by D7-11 in which two penta-domains intertwine to form a dimeric helical-type coil. Remarkably the ligand free structure of this penta-domain closely matches the IGF2 bound state suggesting this may be an intrinsically stable conformation at neutral pH. An interdomain cluster of histidine and proline residues identified between several pairs of domains impart rigidity and may play a role in structural transitions of the receptor at low pH.

Keywords: X-ray crystallography; CI-MPR; IGF2R; P-type lectin; Domain 9; Mannose 6-phosphate.

Introduction:

Protein glycosylation is a ubiquitous and complex set of post-translational modifications that affects protein solubility, function, signalling, dynamics, half-life and trafficking (Moremen et al., 2012). Most of these interactions are mediated by often multivalent binding with a complex family of carbohydrate binding proteins, the lectins (Taylor and Drickamer, 2019). P-type lectins are unique in their ability to bind phosphorylated sugars, principally mannose 6-phosphate (M6P) (Dahms and Hancock, 2002). The two members of the P-type lectin family are the cation-dependent mannose 6-phosphate receptor (CD-MPR) and the cation-independent mannose 6-phosphate receptor/ insulin-like growth factor 2 receptor (CI-MPR/ IGF2R). They both play a critical role in intracellular protein trafficking and lysosome biogenesis (Kim et al., 2009). Together these P-type lectins are responsible for the trafficking of approximately 50 M6P-tagged lysosomal acid hydrolases from the trans golgi network (TGN) to the lysosome (**Figure S1A**). Both the CI-MPR and CD-MPR bind M6P mono-esters (**Figure S1B**) but the CI-MPR is distinct from the CD-MPR primarily in its ability to bind an array of ligands at neutral pH including M6P di-esters as well as non-glycosylated ligands such as IGF2 and CREG (Dahms and Hancock, 2002; Ghobrial et al., 2018; Sacher et al., 2005; Song et al., 2009; Zaccheo et al., 2006).

Loss of CI-MPR binding activity may lead to decreased clearance of proteinases such as cathepsins that are involved in extracellular matrix degradation and promote tumour invasion (Probst et al., 2009; Souza et al., 1997). CI-MPR loss of function is also implicated in lysosomal storage disorders that encompass a broad range of pathologies including neurodegeneration and retarded development in infants and the CI-MPR gene is therefore a target for enzyme/gene replacement therapy (Platt et al., 2012). Loss of IGF2 binding by the CI-MPR can also lead to increased bio-availability of IGF2 which promotes cell growth, proliferation and angiogenesis (Hughes et al., 2019). The CI-MPR therefore acts as a natural tumour suppressor protein with over-expression shown to contribute to increased apoptosis in tumours (Harper et al., 2006).

The CI-MPR is an approximately 300 kDa type I transmembrane glycoprotein (Brown et al., 2002). Ligand binding is attributed to 15 homologous domains (each 124-192 amino acids) in the extracellular region that have between 14-28 % sequence identity to each other and the extracellular domain of the CD-MPR (Brown et al., 2002). Domains 3, 5, 9 and 15 are M6P receptor homology (MRH) domains capable of binding M6P tagged glycoproteins. Domains 1-3 act co-operatively to increase the binding of domain 3 to M6P monoesters by 1000-fold (K_D 0.5 nM versus 500 nM) with domains 1 and 2 stabilising loops of domain 3 involved in carbohydrate binding (Olson et al., 2010). Domain 5 is selective for M6P di-esters and, like domain 3, binding is enhanced by the presence of neighbouring domains (Bohnsack et al., 2009). On the other hand, domain 9 is specific for M6P mono-esters and acts independently from other domains of the CI-MPR (Bohnsack et al., 2009). Domain 15 exhibits low affinity M6P binding with a K_D of 13 and 17 µM, for M6P mono- and di-esters respectively (Olson et al., 2015a). Domain 11 binds the mature,

non-glycosylated form of the potent growth hormone insulin-like growth factor-2 (IGF2, 1-67 amino acids) with nanomolar affinity (K_D ~100 nM) as well as partially processed 'Big' isoforms of IGF2 (1-87, 1-104 and 1-156) (Greenall et al., 2013; Zaccheo et al., 2006). Although the larger glycosylated forms of IGF2 are bound, the interaction is not via M6P but instead mediated by domain 11 which binds IGF2 via a series of conserved hydrophobic and polar interactions (Brown et al., 2002; Williams et al., 2012; Zaccheo et al., 2006). Bound IGF2 is internalised and subsequently degraded (**Figure 1A**) (Brown et al., 2008). Although domain 11 can bind IGF2 independently, a fibronectin insert in domain 13 and residues of domains 6 and 8 increase the affinity for IGF2 (Brown et al., 2008; Devi et al., 1998; Wang et al., 2020; Zaccheo et al., 2006)

To date, high-resolution X-ray or NMR solution structures have been determined for bovine domains 1-3 (PDB codes: 1SZ0, 1SYO, 1Q25 (Olson et al., 2004b, 2004a)), bovine domain 5 (PDB codes: 2KVA, 2KVB (Olson et al., 2010)), human, echidna, chicken and opossum domain 11 (PDB codes: 1GP0, 2LLA, 2L21, 2L2G (Brown et al., 2002; Williams et al., 2012), human domains 11-12 (PDB codes: 2V5N (Brown et al., 2008)) and human domains 11-14 (PDB codes: 2V50 (Brown et al., 2008)) as well as domain 11 variants engineered for high-affinity ligand binding (Frago et al., 2016) (**Figure 1**). Most recently, a low-resolution cryo-EM structure of bovine CI-MPR has been determined at pH 4.5 and at pH 7.4 (domains 4-14) in complex with IGF2 (Wang et al., 2020).

Seeking to understand the innate, cell surface structure of CI-MPR (i.e. the oligomeric state near neutral pH on the cell surface and how this is affected/ regulated by extracellular ligands), we have studied interdomain orientations and their impact on ligand binding, focusing on the area of poor understanding surrounding the selective, high-affinity and independent M6P binding domain 9. Thus, using x-ray crystallography, we report the first high-resolution structure of the elusive domain 9 captured within a rigid multi-domain construct formed by domains 9 and 10 (D9-10). Further we present a structure of human CI-MPR domains 7 to 11 (D7-11) that assembles into a dimeric hub at neutral pH in which the structure of the D9-10 interface is preserved. Alongside a model for the extracellular region of the human CI-MPR, these structures provide insight into two key domain interfaces of CI-MPR that are involved in the switch to an extended multi-domain arrangement at neutral pH.

Results and Discussion:

Expression and purification of domain 9 constructs.

Despite the successful preparation of recombinant CI-MPR domains from birds, monotremes, marsupials and mammals for functional and structural studies (Brown et al., 2008, 2002; Olson et al., 2015b, 2015a, 2010, 2004a, 2004b; Williams et al., 2012), the high-resolution structure of human domain 9 has surprisingly not been reported. At the outset of this work, human CI-MPR domain 9 (residues 1221-1365) was expressed in *E. coli* and formed inclusion bodies (**see supplemental methods**). A range of refolding

protocols yielded soluble protein but in no case was a well dispersed, native-like ^{1}H or ^{1}H-^{15}N HSQC NMR spectrum observed that would indicate correctly refolded protein (**Figure S2A**). *E. coli* expression of domain 9 from three further species (*Bos taurus*, *Lemur catta*, *Ornithorhynchus anatinus*) yielded similar results (data not shown). The CI-MPR has a total of nineteen predicted glycosylation sites, although structural studies have revealed at least two of these are not modified (Brown et al., 2008, 2002). Expression of glycosylated D9 (N1246 and N1312) in HEK cells produced soluble, disulphide bridged, glycosylated protein (**Figure S2B**) but again this yielded poor ^{1}H NMR spectra suggesting this construct was not suitable for structural studies. *In vitro* de-glycosylation of the protein failed to improve the NMR spectra (**Figure S2C**).

Expression and purification of human domains 9-10.

Based on the observation that a human domain 9 protein could not be isolated alone we sought to express an adjacent domain which might generate a stable tandem domain structure. A sequence corresponding to human CI-MPR 9-10 di-domain (D9-10, **Table S1**) was selected for expression in insect cells and was chosen, over an 8-9 di-domain, based on the lack of N-linked glycosylation sites in domain 10 that might therefore minimise sample heterogeneity and assist with subsequent biophysical and structural analysis. D9-10 was expressed in *Sf21* insect cells using EMBacY Multi-Bac (Bieniossek et al., 2008). Recombinant protein expression and optimal harvest time was assessed by SDS-PAGE analysis and YFP emission (**Figure S2D, S2E**). The presence of an N-terminal signal sequence allowed secretion of D9-10 which was subsequently harvested from culture media by Ni-NTA affinity chromatography and purified to greater than 90 % purity with a yield of 7-13 mg/L (**Figure S2F-H**).

The ESI-MS m/z envelope (**Figure S3A**) revealed the presence of multiple species with masses exceeding full length D9-10 (33,587 Da) and consistent with a mixture of typical insect cell derived glycan modifications in which both glycosylation sites of domain 9 were modified (**Figure S3D**) (Aebi, 2013; Aebi et al., 2010; Shi and Jarvis, 2007). Batch to batch preparations showed slight variability in the distribution of glycans (**Figure S3B**) and treatment of D9-10 with PNGaseF confirmed complete loss of glycosylation (**Figure S3C**). D9-10 was functional and demonstrated M6P binding by NMR but showed no response to glucose or mannose as expected (**Figure S4**).

Crystallisation and structure determination of D9-10.

D9-10 crystallised from a solution of 0.1 M SPG (succinic acid, sodium dihydrogen phosphate monohydrate, glycine), 25 % PEG 1500 at pH 9 for X-ray structure determination. The construct crystallised in space group $P2_12_12_1$ with a single molecule in the asymmetric unit. Molecular replacement using homology models of domains 9 and 10 based on reported CI-MPR domains yielded a model that refined to 1.5 Å (Rwork and Rfree values of 19.9 % and 22.8 % respectively) with 98.6 % of the backbone dihedral angles in favoured regions of the Ramachandran plot (**Supplemental Information Table 2**).

The structure of D9-10 (**Figure 2**) is formed from two well-defined domains of similar size, comprised of domain 9 residues D1225-D1365 and domain 10 residues L1366-K1509 and is the first reported to encompass the specific mannose 6-phosphate (M6P) binding site of domain 9. There was no electron density for amino acids V1222-G1224 at the N-terminus of domain 9 or S1510 at the C-terminus of domain 10. Electron density was observed for one N-linked glycan on N1312 of domain 9 (**Figure 2A**) which is positioned through association with domain 9 of a symmetry related partner (**Figure 2B**). Modelling $GlcNAc_2Man_4$ yielded the best fit to the electron density (**Figure 2B inset**). The second glycan predicted to be linked to N1246 was not observed. Both domain 9 and domain 10 form a core β-barrel structure stabilised with four disulphide bonds as previously observed for domains 1-3, domain 5 and domains 11-14 of the CI-MPR (Brown et al., 2008; Olson et al., 2010, 2004a). Two antiparallel β-sheets (βA-D and βE–I) form a flattened, nine-stranded β-barrel linked by four loop regions (termed AB, CD, FG and HI) which vary in length between four and seven amino acids.

Interestingly, the interface between domain 9 and 10 buries surface areas of ~ 600 $Å^2$ (8 %) and 625 $Å^2$ (8 %) respectively of the domain 9 and 10 total surface areas (PISA, Krissinel and Henrick, 2007). The electron density observed in this region was well defined. The interface is stabilised by a combination of hydrophobic interactions, hydrogen bonding and the disulphide bridge between C1333-C1361 at the base of D9 βG and the linker region (**Figure 2C**). The histidine (H1234) in the β-hairpin at the N-terminus of domain 9 sits within a pocket formed by the proline rich BC loop of domain 10 (**Figure 2C**) and is sandwiched between P1422 on domain 10 and P1362 on domain 9, with separations consistent with the formation of CH-π interactions (Nishio et al., 2014; Saha et al., 2007). R1233 also appears to pack against P1419 yielding a CH-π interaction. Y1387 of domain 10 also interacts with F1364 and R1335 to stabilise the opposing side of the interface. Hydrogen bonded residues include residues at the base of domain 9 and residues in the BC loop of domain 10 (N1306-G1416, R1356-S1388, R1233-G1416/Q1414, D1365-S1410) (**Figure 2D**).

Domain 9 M6P binding site and comparisons to other MRH domains.

Domain 9 adopts the conserved MRH fold observed for CI-MPR domains 3 and 5, and the CD-MPR (Brown et al., 2008; Olson et al., 2010, 2002). Sequence alignments and mutagenesis studies have shown that these domains contain a four residue 'QREY' binding site motif present as Q1283, R1325, E1345 and Y1351 in human domain 9 (**Figure 3A**) (Hancock et al., 2002). G1283 on βC forms a hydrogen bond with the 2'OH of terminal mannose, while E1345 and Y1351 on βH and βI respectively can hydrogen bond to the 3'OH and 4'OH. From the FG loop, the charged guanidinium group of R1325 interacts with the 6'OH of the terminal mannose group. H1320, previously identified as essential for M6P binding but

not present in domain 3 (Marron-Terada et al., 2000), also faces into the binding site of domain 9 and engages the 6'OH of the same mannose group.

X-ray structures of bovine domains 1-3 also revealed occupancy of the M6P binding site either by M6P or a non-phosphorylated branched glycan ligand that varied with protein preparation (PDB codes 1Q25 and 1SZ0/1SY0 respectively) (Olson et al., 2004b, 2004a). A comparison between D9-10 and D1-3 structures reveals a common orientation of the glycan and M6P (**Figure 3A**) together with the amino acid residues shown to be essential for M6P binding (Hancock et al., 2002). H1320 in domain 9 bifurcates the positions of the equivalent S326 and S327 residues in domain 3 and would be able to form a direct favourable charge-charge interaction with the phosphate group of M6P in the equivalent pose to the glycan (**Figure 3A**). The orientation of the histidine sidechain is homologous to histidine residue 1291 from the bovine CD-MPR high-resolution structure that binds the phosphate group of M6P (PDB code: 2RL8) (Olson et al., 2008) as well as close structural conservation of other key residues in the M6P binding site (**Figure 3B**).

The interaction between the FG and HI loops defines an important region of the sugar binding pocket across the MRH domains. In domain 9, H1320 lies adjacent to the disulphide bridge (C1319-C1349) that connects the FG and HI loops and together these residues occlude one side of the binding pocket. This region of the binding pocket was, however, proposed to be important for the recognition of M6P di-esters by bovine domain 5 and the accommodation of an additional GlcNAc moiety of this glycan (**Figure 3E**) (Olson et al., 2010). The FG and HI loops of domain 5 are not connected by a disulphide bridge and they lack bulky residues, creating an open binding pocket that can easily accommodate the GlcNAc residue of the M6P diester. Attempting to model in an equivalent diester into the domain 9 structure creates clear steric clashes (**Figure 3F**). This closely resembles the deep narrow pocket that is formed by the surface of the CD-MPR M6P specific binding pocket (**Figure 3C**) which is defined by an almost identical packing of H105, R111 and an equivalent disulphide bridge connecting the FG and HI loops (**Figure 3C**). Although lacking the bulkier histidine residue, the disulphide bridge also partially occludes the binding pocket in domain 3 (**Figure 3D**). Taken together, occlusion of this region of the MRH binding pocket appears to be a common mechanism for ensuring specificity for M6P mono-esters.

Both domains 9 and 3 also show close structural homology to OS-9 (RMSD of 1.3 Å and 4.6 Å respectively), a protein that recognises two α1-6 linked mannose residues on the C-arm of high-mannose type glycans on ER-associated degradation substrates (Satoh et al., 2007). Q1283, R1325, E1345 and Y1351 all have conserved counterparts in OS9 that form interactions with the hydroxyl groups of the bound mannose subunits. The exceptions are Y2155 which is substituted as part of a crucial di-tryptophan glycan binding motif and H1320 which is substituted with D182 that binds the 6'OH of Man(B) (Satoh

et al., 2010). The switch therefore to a polar residue (domain 3 S386/S387) or positively charged residue (domain 9 H1320) appears to be associated with the presence of the negatively charged M6P moiety.

Structure of domains 7-11.

With the structure of D9-10 in hand, we sought to prepare a larger receptor fragment that encompassed this smaller di-domain component. A bovine construct comprised of domains 7-11 (D7-11) has previously been shown to be stable to proteolytic digestion (Westlund et al., 1991) and expressible as a soluble protein that maintains M6P binding (K_d = 0.5 nM) as well as an intact IGF2 binding site (domain 11) (Marron-Terada et al., 2000). In parallel expression experiments, N-terminally His$_6$-tagged human D7-11 was expressed in mammalian HEK293S cells as well as shorter D8-9, D9-10 and D9-11 constructs. For SPR, these multi-domain constructs were also expressed in HEK293T cells with an N-terminal *E. coli* biotin ligase (BirA) acceptor tag and assayed for M6P-binding by SPR using the known CI-MPR ligand Leukaemia inhibitory factor (LIF) (**Figure S5**). The binding affinities (K_D) determined by SPR of each construct were between 60-80 nM (**Figure S5**) and in line with literature values for a single domain 9 construct binding M6P mono-ester glycoprotein GAA (K_D 95 ± 12 nM) (Chavez et al., 2007). The binding affinity of domain 9 does not change significantly between constructs, confirming the unique property of domain 9 that M6P mono-ester binding affinity is independent of the presence of neighbouring domains (Bohnsack et al., 2009; Hancock et al., 2002; Olson et al., 2015a).

D7-11 from HEK293S cells crystallised in a solution of 0.1 M MES, 1.6 M MgSO$_4$, 10 mM M6P at pH 6.5. Initial x-ray crystallographic analysis indicated that the D7-11 crystals were of space group P4$_1$2$_1$2 with two molecules in the asymmetric unit.

To assist in the structure determination of D7-11 by molecular replacement, a construct encoding single domain 8 was expressed in *Sf21* insect cells (as for D9-10) and crystallised from a solution of 1.5 M ammonium sulphate, 0.1 M Tris pH 8.5 and 12 % glycerol for x-ray structure determination. These protein crystals were of space group P 12$_1$1 with two molecules in the asymmetric unit. Molecular replacement using domain 10 from the D9-10 crystal structure yielded a model for domain 8 that refined to 2.6 Å (Rwork and Rfree values of 26.9 % and 31.0 % respectively) with 99.3 % of the backbone dihedral angles in allowed and favoured regions of the Ramachandran plot (**Table S2**). Electron density was observed for V1082-V1220 of domain 8 and S1081 from the N-terminal signal sequence. Although ESI-Mass spectrometry revealed domain 8 is modified by the addition of an N-linked glycan at N1164 (**Figure S2E**), no electron density for the glycan was observed.

Molecular replacement using the X-ray structures of domain 8 (reported here, PDB code: 6Z31), D9-10 (reported here, PDB code: 6Z30) and domain 11 (PDB code: 1GP3) (Brown et al., 2002) yielded a model of D7-11 that refined to 3.5 Å (Rwork and Rfree values of 26.1 % and 30.0 % respectively) with 99.6 %

of the backbone dihedral angles in the allowed and favoured regions of the Ramachandran plot (**Table S2**). There was no electron density for T1359-A1365 at the N-terminus of domain 7 or E1510-T1513 at the C-terminus of domain 11.

The structure of D7-11 (**Figure 4**) comprises residues C934 to C1644 of the human CI-MPR and is the largest fragment of the human CI-MPR solved by x-ray crystallography to date. The asymmetric unit of the crystal contains two 80 kDa copies of D7-11 that associate to form an intertwined 160 kDa homodimer (referred to as chains a and b herein). Each D7-11 polypeptide chain is comprised of five well-defined domains of similar size: domain 7 (C934-P1081, 147 amino acids), domain 8 (V1082-R1221, 139 amino acids), domain 9 (V1222-D1365, 143 amino acids), domain 10 (L1366-S1510, 144 amino acids) and domain 11 (N1511-C1644, 143 amino acids). Both chains wrap around each other primarily through associations between domains 8-10. Domains 7, 8, 9 and 10, which have not previously been structurally characterized, are well defined in the complex and each has a β-barrel core topology composed of nine anti-parallel β-strands. The IGF2 binding domain 11 is less well defined in the complex and associates with domain 10 of the other chain (i.e. D11a to D10b) via the βE-βI sheet. This is the same region of domain 11 observed to pack against domain 12 (Brown et al., 2008) and may therefore be an artefact arising from the burial of this hydrophobic face of domain 11, possibly due to the absence of domain 12. This packing may be sub-optimal and may explain the poor electron density observed for this domain.

Domain 7 lacks an identified ligand but the AB, CD, FG and HI loops point away from the dimeric hub. Domain 7 has a short AB loop, extended CD and HI loops and R1065 (N-terminal of βI) substitutes for disulphide bridge to the FG loop. Together with residues K963 (AB loop), K1000, R1003 (CD loop) and K1030 (FG loop), these residues form a pocket of positively charged residues (**Figure 5A**). However, compared to domains 3, 9 and 11 which have extended AB loops but shorter CD loops, the domain 7 loops do not form an obvious binding groove.

The dimeric interface of D7-11 is formed from predominantly hydrophobic contacts with a total buried surface area of approximately 19500 Å^2 and a solvent accessible area of 66400 Å^2 (PISA (Krissinel and Henrick, 2007)). The binding site loops of domain 9 are solvent exposed and, as observed in the D9-10 structure, a mannose residue from the modelled portion of the N1312-linked glycan of domain 9 extends into the M6P binding site. However, in the D7-11 dimer this is a symmetric interaction with both N-linked glycans binding to opposing D9 binding sites in the dimer (**Figure 5B**). The terminal mannose residues are oriented to bring the 6'OH within ~ 3 and 6 Å of H1320 in the two sites and the two interactions bring the FG loops of neighbouring domain 9 molecules into close proximity. The conserved QREY residues of the domain 9 M6P binding site interact with the terminal mannose residue supplemented by additional interactions to the preceding sugar (i.e. 3'OH to Y1255 at the N-terminus of βB). Interestingly, the M6P

binding site of D9 and the IGF2 binding site of D11 point away from one another on opposing surfaces, meaning that in this conformation at least, the binding of either ligand is unlikely to occlude the binding site of the other through steric interaction.

The most solvent exposed domain is domain 7, although this structure lacks the neighbouring domain 6 of the full ectodomain. The N-terminal βN and βN' strands, the GH loop and the C-terminal loop of βI from domain 7 all contact domain 8 of the same chain (i.e. D7a to D8a) (**Figure 5C**). By contrast domain 8 and domain 10 form the most extensive inter-domain contacts and the core of the dimer (**Figure 5D**). Superimposing the single domain 8 crystal structure (PDB code: 6Z31) and domain 8 of the D7-11 structure (PDB code: 6Z32) gives a very close superposition with an RMSD of 0.5 Å and the extensive packing around domain 8 in D7-11 suggests the role of this domain is purely structural. Interestingly however there are no homo-dimeric contacts (i.e. 8a to 8b and 10a to 10b) for either of these domains and instead the domain 9 dimer forms a capstone that binds a ring of domains formed from 8/10-8b/10b.

Interdomain orientation in CI-MPR D1-3, D9-10, D7-11, D11-14 and CD-MPR.

The observation of a well-defined interdomain boundary in the D9-10 structure led us to compare this with the domain arrangement in D7-11. A comparison of the interface between domains 9 and 10 in the D9-10 and D7-11 X-ray structures revealed a close similarity. Superimposing over only the secondary structure elements yielded an RMSD of 3.1 Å for backbone atoms, consistent with this close alignment. Moreover, the inter-domain interactions also appear to be well conserved, with His-Pro inter-domain interactions identified in D9-10 being present in the D7-11 structure (**Figure 6A-B**). Extension to domains 8-10 reveals a compact tri-domain resembling that of D1-3 (**Figure 6C,** PDB codes: 1SZ0, 1SYO, 1Q25). Other tri-domain combinations (D7-9 and D9-11) are arranged in a more elongated or open distributions.

Small angle X-ray scattering (SAXS) on the D9-10 construct was used to further probe this conformation in solution and confirmed a folded flat entity with a calculated molecular weight associated with a dimer (apparent molecular weight 70.5 kDa (as calculated by SAXSMoW 2.0); expected dimeric molecular weight 70.8-71.6 kDa, **Table S3**). *Ab initio* shape envelope modelling of D9-10 was consistent with the D7-11 crystal structure, with symmetric cross-binding of the two N-glycan moieties (**Figure S6**). Fitting required flexibility within the linker region between domains 9 and 10 (residues 1358-1364) but not the principle D9-10 interface and a χ^2 value of 1.68 for a one state model of D9-10 indicated that these models were a good fit to the experimental data. The shape parameters (Rg and Dmax) associated with D9-10 were largely unaffected by the addition of a 100-fold excess of M6P, with the structure remaining dimeric and again fitting to the D9-10 dimer arrangement of D7-11 (χ^2 value of 1.75) (**Figure S6, Table S3**). De-glycosylation, however, resulted in monomerization with an apparent molecular weight of 33.7 kDa (expected monomeric molecular weight 33.2-33.6 kDa). Nonetheless the scattering data could be

modelled to a shape envelope consistent with the D9-10 didomain structure (χ^2 value of 1.31, **Figure S5, Table S3**). It appears therefore that D9-10 forms a rigid body at neutral pH generated from key interactions between these domains.

The SAXS data correlates with the crystal structure of D7-11 in which there is no contact between opposing domain 10s and dimerization is driven by the bridging N-linked glycan at N1312 of D9 rather than a host of specific protein-protein interactions. Conversely, crystal structures of the CD-MPR at pH 6.5 (PDB codes: 2RL8, 2RL9, 1KEO) reveal packing of the two βE-I surfaces against one another, with the HI loop of one chain interacting with the N-terminus of the other chain. Therefore, despite the similarities in the domain 9 and CD-MPR binding sites, their mechanisms of dimerisation are distinct.

We next searched for similar interactions to D9-10 between other CI-MPR domains where high-resolution structures were available. The X-ray structure of a D11-12 didomain at neutral pH (pH 7.5, PDB code: 2V5N) has previously been reported and revealed interactions between the domains formed by a hydrophobic patch of domain 11 formed from βE-βI β-sheet packing against the BC loop of domain 12 (Brown et al., 2002). This interaction was maintained in two further X-ray structures of D11-14 and D11-13 bound to IGF2 (PDB codes: 2V5O and 2V5P respectively) suggesting a degree of rigidity between D11-12 rather like D9-10 (**Figure 6A**). Interestingly this interface also includes similar interdomain interactions, with H1641 of domain 11 interacting with P1697 and P1700 of the proline rich BC loop of domain 12 as well as several other hydrophobic interactions (**Figure 6B**). H1641 resides on βI of domain 11 rather than the loop between βNA' and βNA" observed in D9-10 leading to a different association and side-on packing of the domains 11-12 crossed β-sheets.

Model for human CI-MPR extracellular region.

The initial model for the CI-MPR extracellular region was proposed by Olson et al. (**Figure 7A**) and comprised five tri-domain units (Olson et al., 2004b). This was consistent with the compact X-ray structure of D1-3, D1-3 and D4-6 proteolytic stability and homology modelling the association of the remaining domains (D7-9, D10-12, D13-15) (Olson et al., 2004b; Westlund et al., 1991). This model was later modified by Brown et al. to include the D11-14 X-ray structure and dimerization interfaces from domains 3, 5, 9, 12 and 15 while maintaining the three N-terminal tri-domains D1-9 (**Figure 7B**) (Brown et al., 2008). The physiological oligomeric state of the CI-MPR at neutral pH is poorly understood. For example, mammalian expressed recombinant D1-15 was found to be monomeric by size exclusion chromatography coupled to multi-angle light scattering (Dwyer et al., 2020), while both monomeric and dimeric forms of sCI-MPR purified from bovine livers have been observed by native gel electrophoresis (Byrd et al., 2000). Maintaining D1-3, D7-10 and D11-14 in their dimeric forms (**Figure 7C**) and combining this with the D7-11 dimer produces a model for the CI-MPR extracellular region that is

different to previous proposals (**Figure 7C**). Our model shows an elongated structure with approximate dimensions of 260 by 130 by 92 Å assembled around the 7-11 core (**Figure 7C**). The role of domain 12 in receptor dimerization has been proposed (Brown et al., 2008; Byrd et al., 2000; Kreiling et al., 2005), but the mechanism of dimerization in terms of protein-protein interactions has not been explicitly proven. The assembly of the D7-10 hub and the rigidity of the D11-12 interface suggests minimal interactions between domain 12s but does permit extensive inter-chain contacts between domains 12 and 13. Similarly the positioning of the two domain 7s seems to preclude the incorporation of domain 5-domain 5 contacts suggesting domain 5 may not act as a dimerization point or domain 7-domain 8 contact may be flexible (Olson et al., 2010).

Despite the crowding caused by this coiled dimerization, the model remains compatible with simultaneous multi-ligand binding. For example, the M6P binding site of domain 3 is positioned on the opposite face of the D1-3 tri-domain to the uPAR and plasminogen binding sites at the N-terminus of domain 1 (Leksa et al., 2002; Nykjær et al., 1998; Olson et al., 2004b). Similarly domains 3, 5, 9 and 11 are distributed so that IGF2 and M6P tagged proteins can bind simultaneously (Brown et al., 2008; Olson et al., 2004b). There have also been multiple reports that bi-phosphorylated oligosaccharides exhibit higher affinity binding to the CI-MPR than mono-phosphorylated ligands (Distler et al., 1991; Tong et al., 1989) and a dynamic state of CI-MPR that brings the M6P mono-ester sites of domain 3 and 9 to within 45-85 Å has been proposed (Olson et al., 2004b). The distance between domain 3 and 9 of the same chain in our model is ~58 Å. Conversely our dimeric model separates the M6P binding sites of the two domain 9s by a distance of ~22 Å which might be compatible with binding ligands bearing more than one M6P glycan (Fei et al., 2008).

A recent cryo-EM structure of monomeric bovine CI-MPR domains 4-14 in complex with IGF2 at pH 7.4 (PDB code: 6UM2 (Wang et al., 2020)) (**Supplemental Figure 7**) has also reported a helical twist arrangement of domains. A ligand free model was not solvable at neutral pH but a low resolution cryo-EM structure of bovine CI-MPR at pH 4.5 in the absence of IGF2 (PDB code: 6UM1 (Wang et al., 2020)) (**Supplemental Figure 7**) showed it formed a more compact, collapsed structure (*cf* 170 by 80 by 80 Å versus 230 by 170 by 70 Å). Superimposing one chain of the domain 7-11 dimeric core with corresponding domains of the extended D4-14-IGF2 complex, however, gave an RMSD of only 9.4 Å (**Figure 7D**). Thus, identical packing around D9 can be derived from interaction with one of two distinct ligands, IGF2 or glycan, or, more likely, by neutral pH conditions.

We have identified a conserved interaction at the interface of domains 9-10 which is also present between domains 11-12 (**Figure 6A-B**). These interactions are conserved in the IGF2 bound structure at pH 7.4, with the analogous histidine residues of domain 9 (H1234 human, H1243 bovine) and domain 11 (H1641 human, H1650 bovine) sitting in homologous proline pockets of domains 10 and 12 respectively (**Figure**

7E) and are unique to these pairs of domains. The bovine cryo-EM structure solved at low pH shows that these interfaces are broken and the relative orientations of domains 9-10 and 11-12 are altered significantly, with the histidine now sitting outside the pocket of proline residues in each case (**Figure 7E, 7F**). This again may be driven by titration of the histidine group and unfavourable interactions generated by burial of this positively charged residue at low pH. Together, D9-10 and D11-12 may form a pair of important hinge regions that contribute to a mechanism of receptor rearrangement and ultimately ligand dissociation in the acidic late endosome. In this scenario, low pH may dissociate the M6P ligand from D9 as well as collapse the receptor by disruption of the D9-10 and D11-12 inter-domain interaction, both through protonation of histidine residues in the M6P binding site and His-Pro interface respectively.

Conclusions:

In summary, we report the high-resolution crystal structure of human CI-MPR domains 9-10, which not only confirms the expected β-barrel fold previously observed for other CI-MPR domains and the positioning of the conserved sugar binding residues Q1283, R1325, E1345 and Y1351 in the domain 9 binding site, but also reveals a key His-Pro interdomain interaction. Also observed at the interface of D11-12, this interaction is pH sensitive, suggesting a possible role in ligand dissociation in the acidic environment of the late endosome.

In both the multi-domain constructs reported here, D9-10 and D7-11, domain 9 dimerises through a bridging N-linked glycan with a terminal mannose residue. Dimerisation of D9-10 was unaffected by the addition of 100-fold excess of M6P suggesting the avidity interactions are likely to be stable at neutral pH *in vivo*. Thus, questions arise surrounding the ground state of the CI-MPR at the cell surface, how receptor dimerisation is regulated and how the domain 9 M6P binding site is unmasked to allow ligand binding.

Author contributions:

A.B., C.W., H.H. performed the experiments. A.B. and K.H. collected the X-ray diffraction data. A.B., A.M. and C.W. determined the crystal structures. A.B. and A.W. collected and analysed SAXS data. I.B. provided reagents and supervision for baculovirus/insect cell expressions. A.B. and M.C. wrote the manuscript that was reviewed by all authors. The project was supervised by M.C. and B.H.

Declaration of Interests:

The authors declare no competing interests.
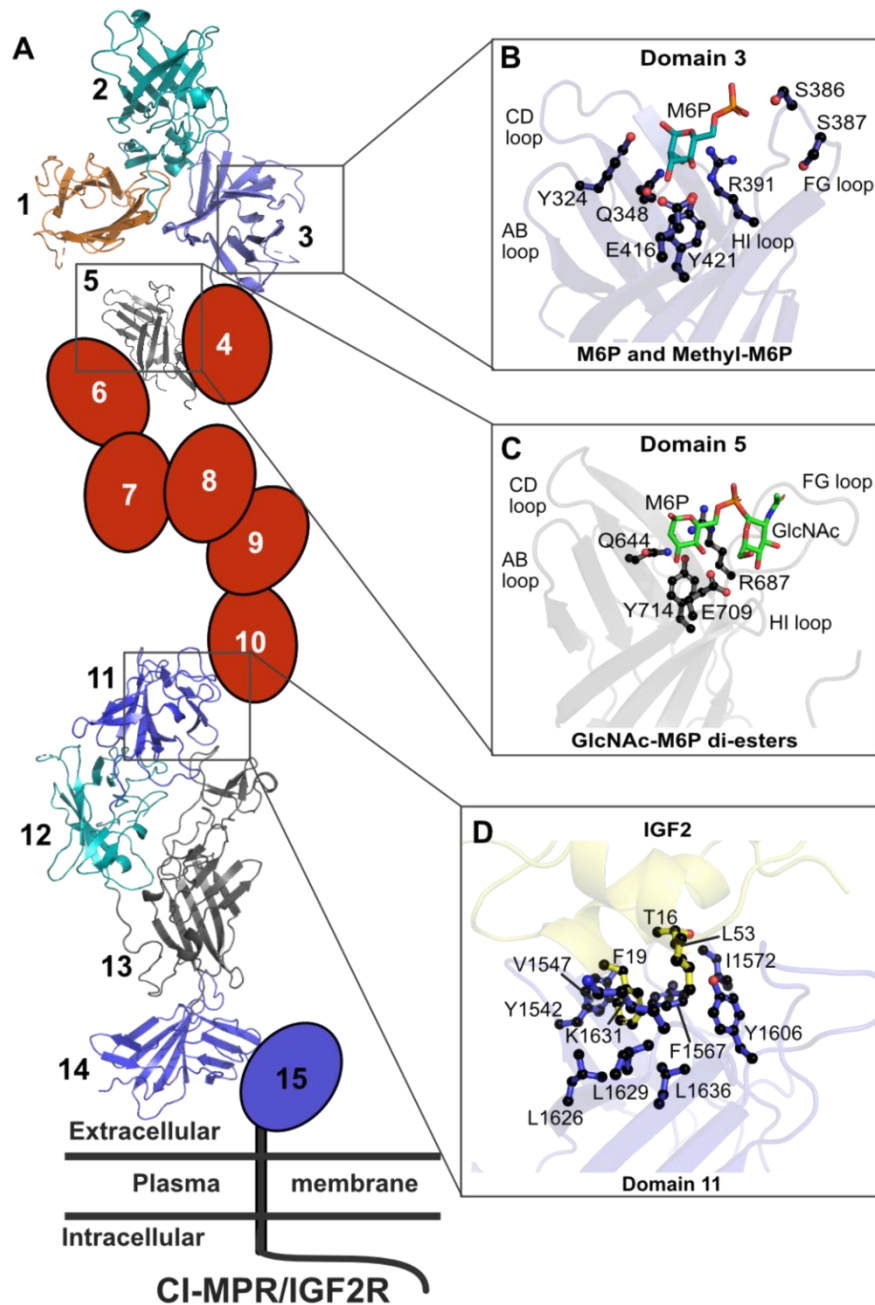
Figure 1: Structural characterisation of CI-MPR binding domains.

A: The extracellular region of the CI-MPR consists of 15 homologous domains. Bovine D1-3 has been crystallised with either bound M6P or mannosylated N-linked glycan (PDB codes: 1SYO/1SZ0 and 1Q25). The structure of bovine D5 with methyl-M6P-GlcNAc bound has been determined by NMR (PDB code: 2KVB). The structure of human D11 has been determined for D11 alone and within the multi-domain constructs D11-12, D11-14 and D11-13 in complex with IGF2 (PDB codes: 2V5N, 2V5O, 2V5P respectively).

B: Inset shows the structure of the D3 binding site with M6P bound (cyan sticks). The conserved sugar binding residues Q348, R391, E416, Y421 and S386 and S387 on the FG loop are shown as sticks.

C: The binding site of D5 (PDB code: 2KVB) with M6P-GlcNAc di-ester modelled (green sticks). The conserved sugar binding residues Q644, R687, E709, Y714 are shown as sticks.

D: The IGF2 binding site of D11. Interacting residues T16, F19 and L53 of IGF2 and Y1542, V1547, F1567, I1572, Y1606, L1626, L1629, K1631 and L1636 of D11 are shown as sticks.

14
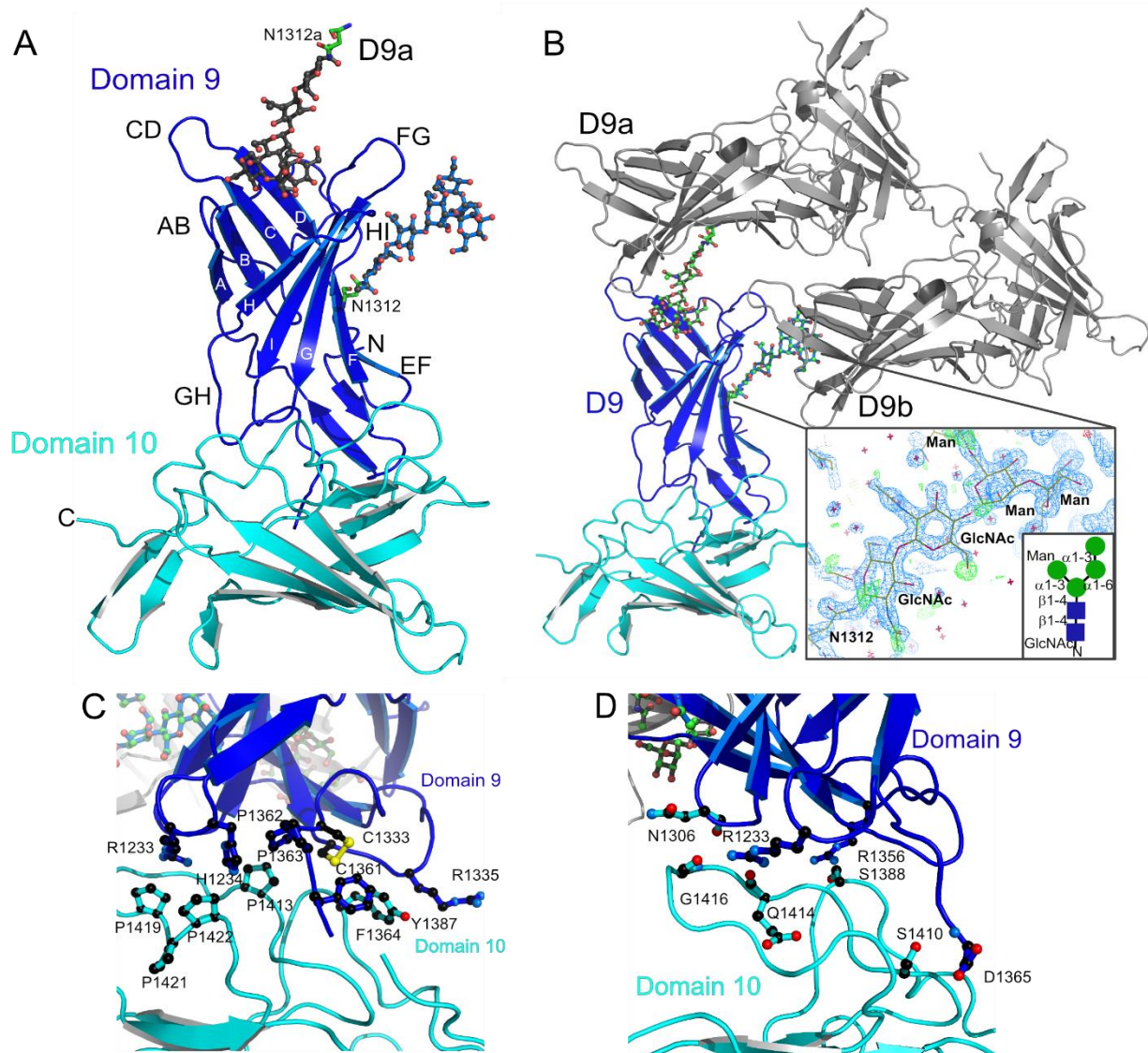
**Figure 2: X-ray crystal structure of human CI-MPR domains 9-10 with N-linked glycan of neighbouring domain 9 bound.**

A: The structure of human CI-MPR domains 9-10 at 1.8 Å resolution. Each domain forms a nine-stranded, flattened β−barrel stabilised by four disulphide bridges (C1227-C1262 (at the N-terminus), C1270-C1282 (in the BC loop and βC), C1319-C1349 (between the FG and HI loops), C1333-C1361 (in βG and the linker region) in domain 9 and C1369-C1408, C1420-C1427, C1461-C1494, C1476-C1506 in domain 10.

B: The structure of domains 9-10 (blue) and two symmetry mates (D9a and D9b in grey) showing the intermolecular glycan bridge. The D9 binding site (blue) is occupied by the GlcNAc$_2$Man$_4$ N-linked glycan from N1312 of D9a (grey). Meanwhile the N-linked glycan of D9 (blue) is bound by D9b (grey). The inset shows the electron density and structure of the bridging glycan (blue squares represent GlcNAc, while green circles represent mannose).

C: H1234 in the β-hairpin at the N-terminus of domain 9 sits within a pocket formed by the proline rich BC loop of domain 10 and is sandwiched between P1422 on domain 10 and P1362 on domain 9 with separations consistent with the formation of CH-π interactions (Nishio et al., 2014; Saha et al., 2007). R1233 (left) also appears to pack against P1419 yielding a CH-π interaction. Y1387 (right) of domain 10 also interacts with F1364 and R1335 of domain 9 to stabilise the opposing side of the interface.

D: Hydrogen bonds form between the base of domain 9 and residues in the BC loop of domain 10 (N1306-G1416, R1356-S1388, R1233-G1416/Q1414, D1365-S1410).
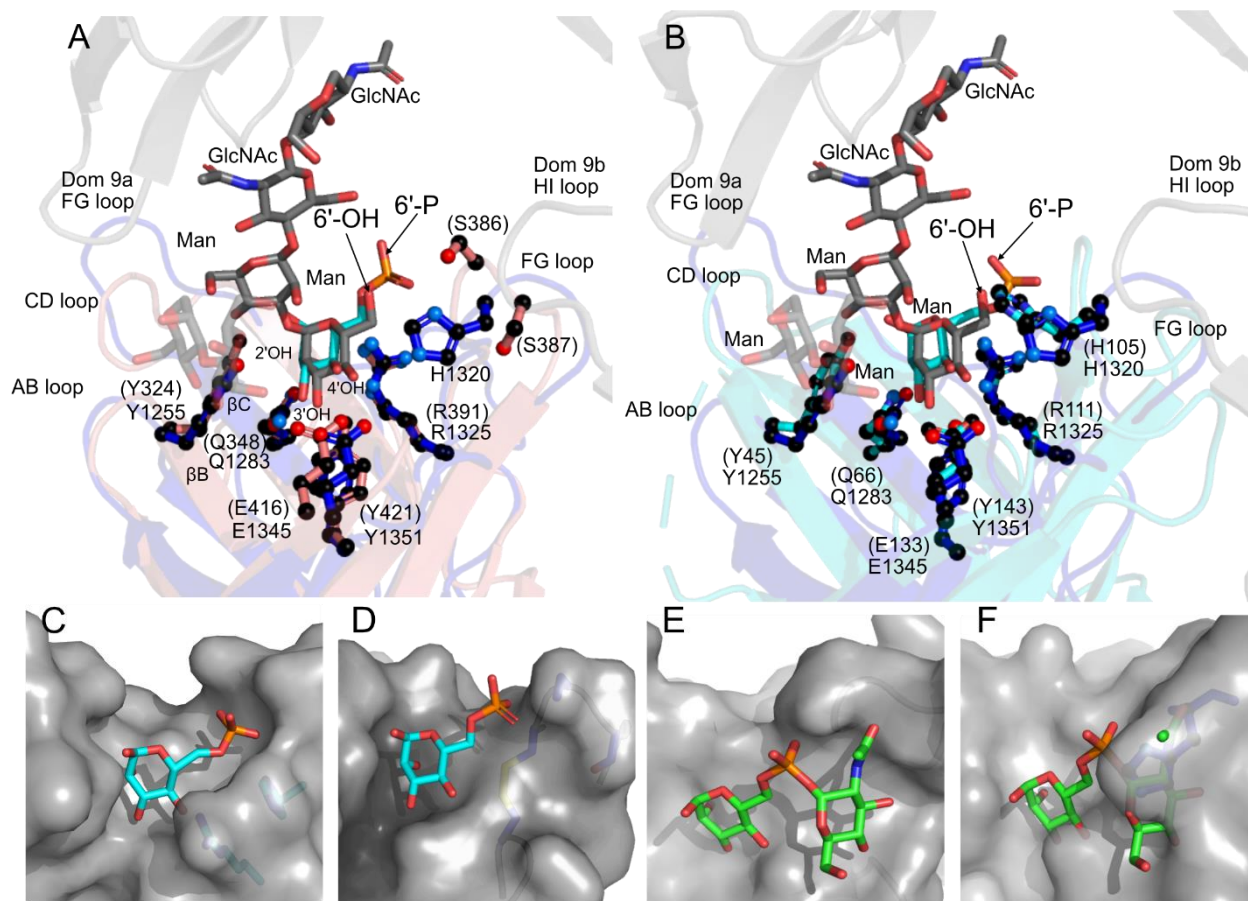
Figure 3: Comparison of the domain 9 binding site with the MRH domains of P-type lectins.

A: Structure of the N-linked glycan (grey sticks) observed in the D9-10 crystal structure (blue cartoon) superimposed with M6P (cyan sticks) observed bound to domain 3 of bovine CI-MPR (orange cartoon, PDB code: 1SZ0). Sugar binding residues of domain 9 are labelled and equivalent domain 3 residue numbers are given in brackets. Q1283 on βC forms a hydrogen bond with the 2'OH of the mannose (not visible in this view), while E1345 and Y1351 on βH and βI respectively can hydrogen bond to the 3'OH and 4'OH. The N-linked glycan branches, with the α1-3 branch of the glycan sitting outside the M6P binding site. Y1255 at the top of βB may form hydrogen bonds with the 4'OH of this branched mannose residue (shown in translucent grey). Homologous residues in the domain 3 binding site are shown as red sticks and labelled in brackets. H1320 in the FG loop of domain 9, which restricts M6P di-ester binding, is replaced by S386 and S387 in the FG loop of domain 3.

B: Analogous comparison of human CI-MPR domain 9 (blue cartoon) with the bovine CD-MPR binding site occupied by M6P (cyan, PDB code: 2RL8). The conserved sugar-binding residues of the CD-MPR are shown as cyan sticks and labelled in brackets. The critical H1320 (H105) adopts a similar juxtaposition to the 6'OH or 6'-P in both structures underlying its critical role in binding M6P mono-esters.

C: The binding surface of the CD-MPR (grey) with bound M6P mono-ester (cyan, PDB code: 2RL8). H105 and R111 are shown as sticks within the surface render.

D: The binding surface of bovine domain 3 (grey) with bound M6P mono-ester (cyan, PDB code: 1SZ0). S386 and S387 of the FG loop and the disulphide bridge between C385 and C419 of the FG and HI loops respectively are shown as sticks.

E: The binding surface of bovine domain 5 (grey, PDB code: 2KVB) with a modelled GlcNAc-M6P di-ester (green sticks) bound in the more open binding pocket. Domain 5 lacks a disulphide bridge between the FG and HI loops and a bulky histidine on the FG loop.

F: The binding surface of domain 9 (grey) with the modelled GlcNAc-M6P di-ester (green sticks). H1320 (blue sticks) on the FG loop occludes di-ester binding.
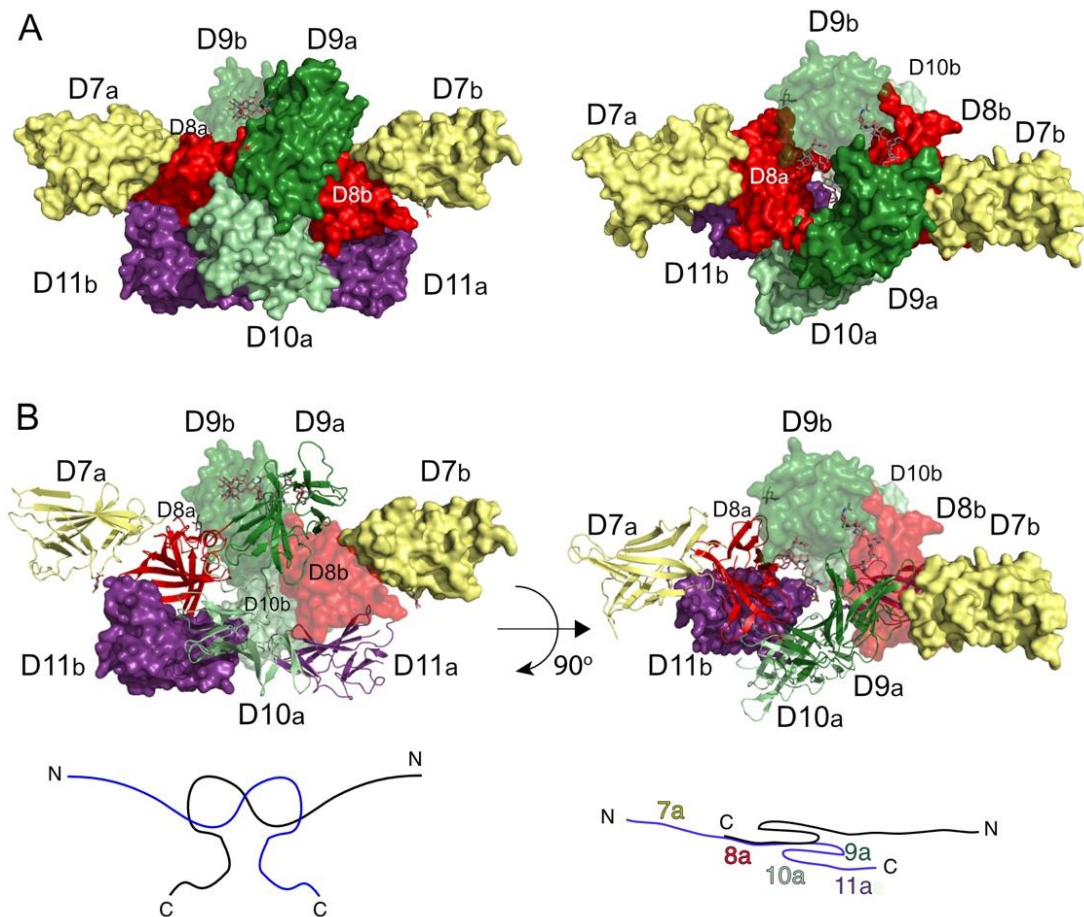
Figure 4: Structure of human CI-MPR domains 7-11 homodimer.
A: Structure of human CI-MPR D7-11 at 3.5 Å resolution. Both monomers are shown in surface representation with the two chains labelled a or b and coloured yellow (domain 7, C934-P1081), red (domain 8, V1082-R1221), dark green (domain 9, V1222-D1365), light green (domain 10, L1366-S1510) and purple (domain 11, N1511-C1644). The structure on the right is rotated by 90° and the glycans extending into the opposing domain 9 M6P binding sites are shown as sticks.
B: One monomer is shown in cartoon format, the other in surface representation and two orientations are shown. Below each surface representation is a trace illustrating the path of the intertwined polypeptide chains of the dimer.
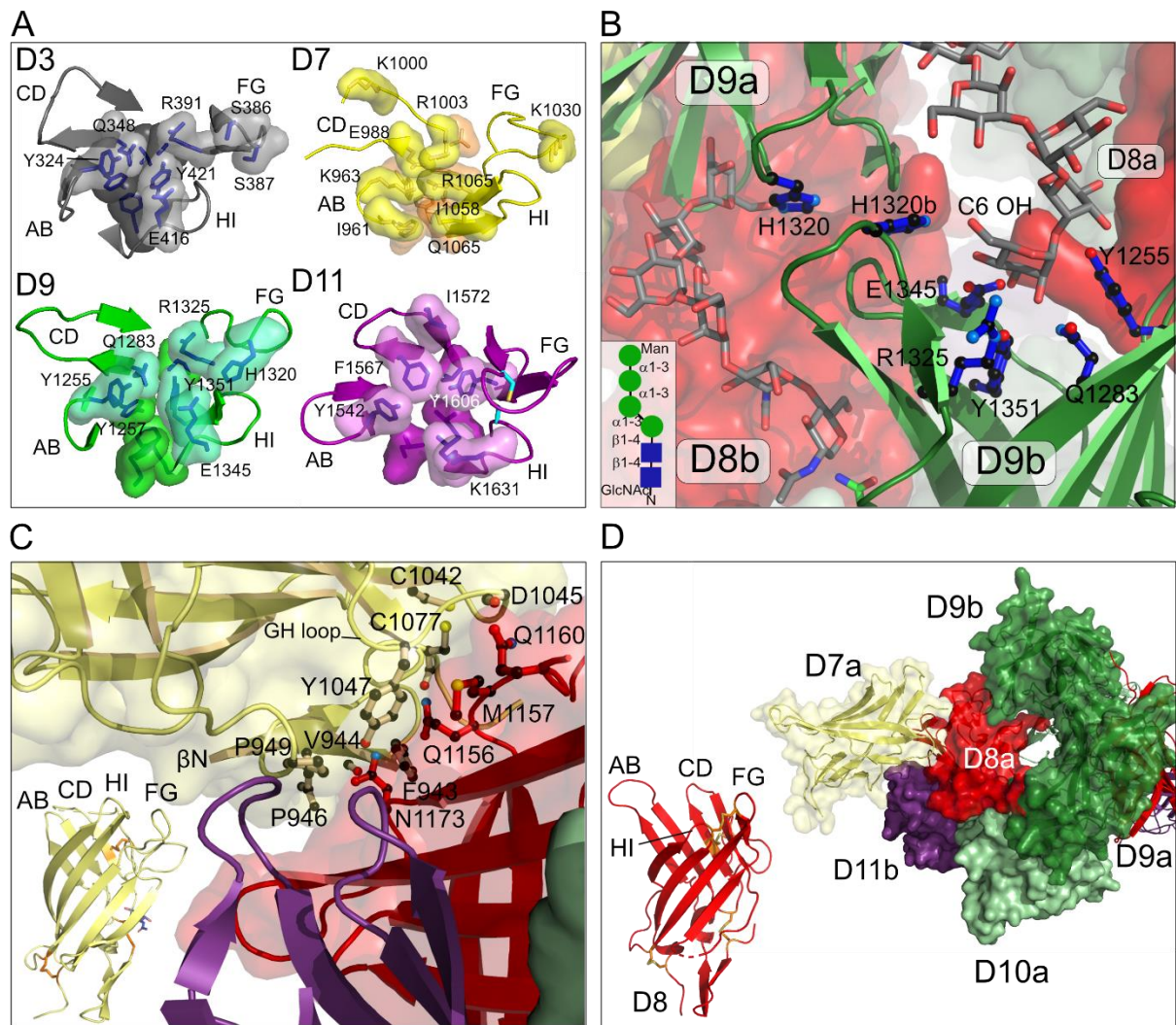
Figure 5: Domain packing and binding sites in D7-11

A: Comparison of domains 3, 9 and 11 binding sites with the comparable site of domain 7. M6P binding residues of domains 3 and 9 and IGF2 binding residues of domain 11 are highlighted. Residues that form the equivalent surface of domain 7 are highlighted, being dominated by positively charged groups.

B: Interface of domain 9 structures. Glycans (grey sticks) extend into the opposing domain 9 M6P binding site. Residues of the conserved QREY motif (Q1283, R1325, E1345, Y1351) and H1320 of the FG loop are show as sticks. Y1255 which binds a mannose group adjacent to the terminal mannose residue is also shown. Inset shows the structure of the bridging N-linked glycan (blue squares represent GlcNAc, green circles represent mannose).

C: Residues at the domain 7-domain 8 interface in the D7-11 crystal structure are shown as sticks. Domain 7 (yellow, bottom left) forms the same β-barrel structure observed for previous CI-MPR domains 1-3, 5 and 11-14 (Brown et al., 2008a; Olson et al., 2004a, 2010).

D: The crystal structure of domain 8 (red, left) was determined alone (PDB code: 6Z31) and within the D7-11 structure (PDB code: 6Z32). Within D7-11, domain 8 is a central core domain that interacts with each of domains 7, 9, 10 and 11 but not with D8b.

18
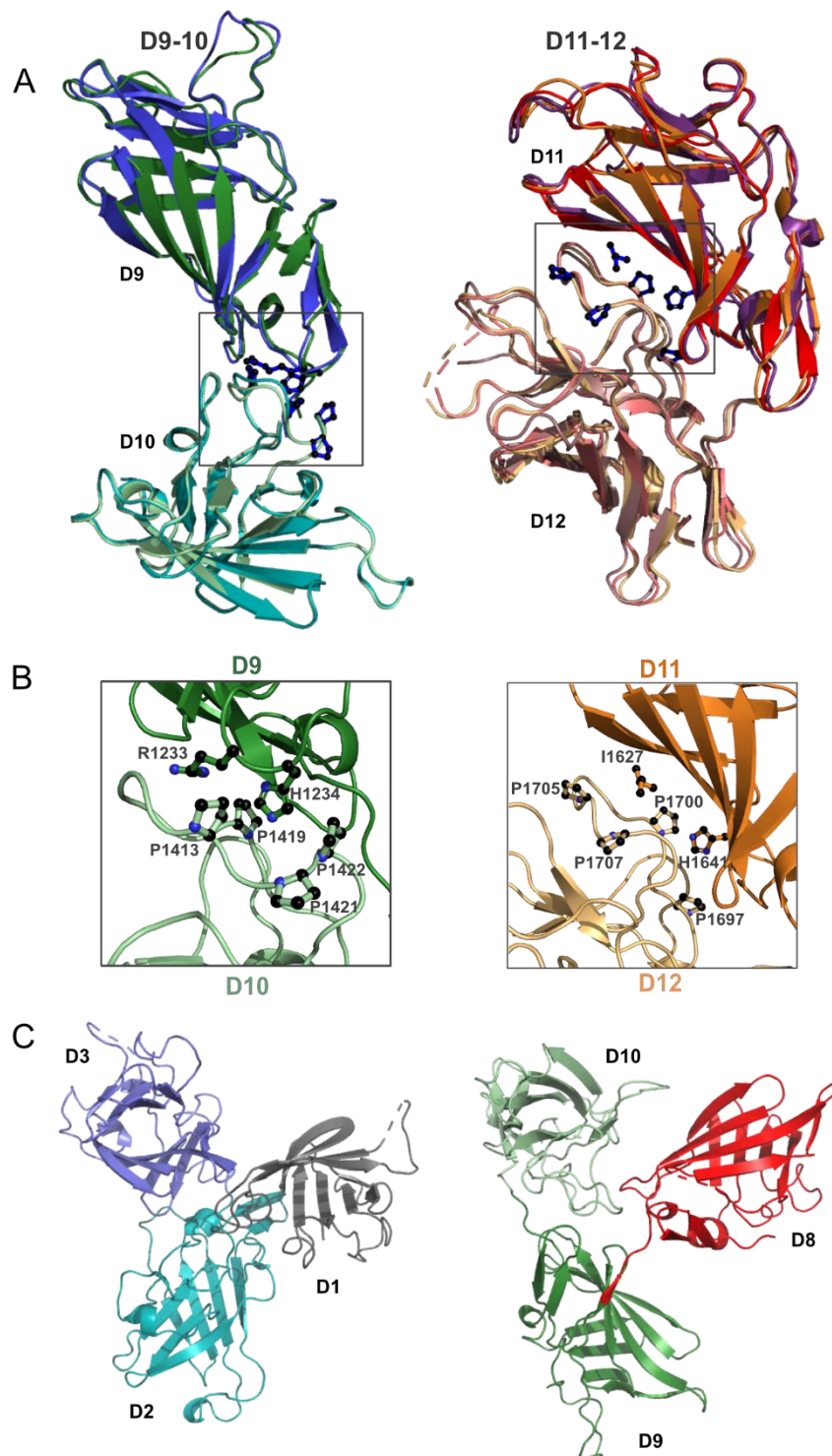
Figure 6: Structure of didomain units 9-10 and 11-12 and relation to higher order structures.

A: Domains 9 and 10 are in the same orientation in the D9-10 structure (blue) and D7-11 structure (green). Domains 11 and 12 (shown) also form a rigid unit that is conserved in the crystal structures of D11-12 (purple, PDB code: 2V5N), D11-13 (red, 2V5P) and D11-14 (orange, 2V5O).

B: H1234 of D9 sits in a pocket of P1413, P1419, P1421 and P1422 on the BC loop of D10. A similar His-Pro interaction to that in D9-10 is seen at the D11-12 interface with H1641 and I1627 of D11 forming CH-π interactions with P1697, P1700, P1705 and P1707 of D12 (PDB codes: 6Z32 and 2V5O respectively).

C: Domains 8-10 of the D7-11 structure form a compact tri-domain orientation similar to that observed for D1-3 (PDB code: 1SZ0).
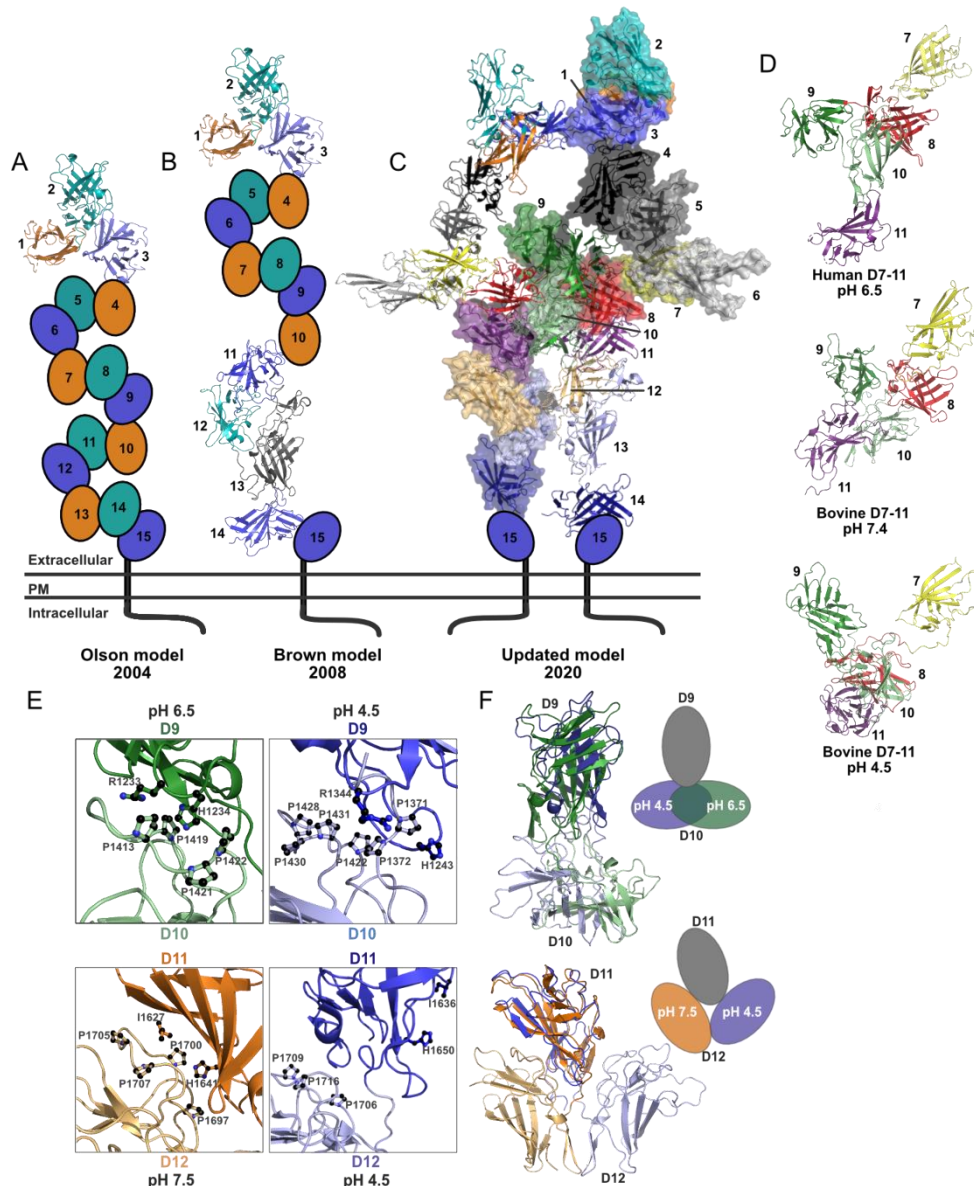
Figure 7: Interdomain orientation of the extracellular region of the CI-MPR.

A: The initial model of CI-MPR domains 1-15 proposed by Olson et al. in 2004 consists of five compact tri-domains (D1-3, D4-6, D7-9, D10-12, D13-15) based upon the crystal structure of domains 1-3 (cartoon, PDB code: 1SYO/1SZ0) (Olson et al., 2004b).

B: The model of CI-MPR domains 1-15 proposed by Brown et al. in 2008 kept the tri-domain arrangement for domains 1-9 but updated the lower half of the receptor based upon the crystal structure of domains 11-14 (cartoon, PDB code: 2V5O), which formed an elongated arrangement (Brown et al., 2008).

C: Our updated model of domains 1-15 is based upon the crystal structure of human CI-MPR domains 7-11 (PDB code: 6Z32). Crystal structures of domains 1-3, domains 11-14 and solution NMR structure of domain 5 were also incorporated (PDB codes: 1SZ0, 2V5O, 2KVA respectively). Domain 15 was not modelled due to its interaction with the transmembrane region and membrane.

D: A single chain of human D7-11 determined by X-ray crystallography at pH 6.5 (PDB code: 6Z32) forms a similar arrangement to D7-11 of the recent cryo-EM structure of bovine CI-MPR domains 4-14 determined in complex with IGF2 at pH 7.4 (PDB code: 6UM2). Superimposition over all atoms gives an RMSD of 9.4 Å. Meanwhile, at pH 4.5 bovine D7-11 (PDB code: 6UM1) forms a more compact structure.

E: Panels showing the His-Pro pockets at the D9-10 and D7-11 interfaces. At neutral pH, in the structures of human D7-11 (green) pH 6.5 and human D11-12 pH 7.5 (orange, PDB code: 2V5N), histidine sits within a pocket of proline residues. However, at low pH 4.5, (bovine D1-14 pH 4.5, blue, PDB code: 6UM1) this interaction is disrupted.

F: Superimposition of D9-10 of the human D7-11 crystal structure (green, pH 6.5) with D9-10 of bovine D1-14 structure (blue, pH 4.5, PDB code: 6UM1) and of D11-12 of human D11-12 crystal structure (orange, pH 7.5, PDB code: 2V5N) and bovine D1-14 structure (blue, pH 4.5, PDB code: 6UM1) reveals the change in domain arrangement upon lowering pH.

STAR Methods:

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
Sf21 insect cells	ThermoFisher Scientific	Cat # 11497013
HEK293S mammalian cells	ATCC	Cat # 3022
HEK293T mammalian cells	ATCC	Cat # 3216
Chemicals, Peptides, and Recombinant Proteins		
Mannose 6-phosphate monosodium salt	Insight Biotechnology	Cat # 70442-25-0
ESF 921 media	Expression Systems	Cat # 96-001-01
Dulbecco's Modified Eagle Medium (DMEM)	Gibco	Cat # 11966025
Fetal bovine serum	Gibco	Cat # 12676029
Restriction endonucleases: EcoRI, HindIII, MfeI AvrII,	ThermoFisher Scientific	Cat # FD0274, FD0504, FD0754, FD1564
Human CI-MPR domain 8 protein	This work	N/A
Human CI-MPR domains 9-10 protein	This work	N/A
Human CI-MPR domains 7-11 protein	This work	N/A
mCherry-LIF-N1 protein	This work	N/A
Oligonucleotides		
Domain 8 forwards primer: GTG GAC TGC AAA GTT ACT GAC	This work/ IDT	N/A
Domani 8 reverse primer: ACG CAC GAC AGG GCA	This work/ IDT	N/A
Domain 9 forwards primer: GTT GAA GGC GAC AAC TGC GAA GTG	This work/ IDT	N/A
Domain 10 reverse primer: GGA CTT CAT AGG GCA AGC AGT	This work/ IDT	N/A
Deposited Data		
Human CI-MPR domain 8 structure	This work	6Z31
Human CI-MPR domains 9-10 structure	This work	6Z30
Human CI-MPR domains 7-11 structure	This work	6Z32
Human CI-MPR domains 9-10 SAXS data	This work	SASDH59,69,79, SASDJ23
Software and Algorithms		
ATSAS (including DAMMIF, DAMMAVER, DAMMIN, SUPCOMB)	(Franke et al., 2017)	https://www.embl-hamburg.de/biosaxs/software.html
CCP4 (including Privateer, REFMAC5 and ProSMART)	(Winn et al., 2011)	http://www.ccp4.ac.uk/
Coot	(Emsley et al., 2010)	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/Coot/
FoXS (including MultiFoXS)	(Schneidman-Duhovny D, Hammel M, Tainer JA, 2016, 2013)	https://modbase.compbio.ucsf.edu/foxs/
Mestrenova	-	https://mestrelab.com/download/mnova/
PISA	(Krissinel and Henrick, 2007)	https://www.ebi.ac.uk/pdbe/pisa/
PHENIX (including Phaser, Autobuild and phenix refine)	(Adams et al., 2010)	https://www.phenix-online.org/
PyMOL	Schrodinger, LLC	https://pymol.org/2/
ScÅtter	(Rambo, n.d.)	https://bl1231.als.lbl.gov/scatter/

TLSMD server	(Painter and Merritt, 2006a, 2006b)	http://skuld.bmsc.washington.edu/~tlsmd/
Xia2 (including XDS and SCALA)	(Winter, 2010)	https://xia2.github.io/index.html
Other		
HisPur Ni-NTA resin	ThermoFisher Scientific	Cat # 88221
Superdex 75 10/300 column	GE Healthcare	Cat # 17517401
ResourceQ 1 mL column	GE Healthcare	Cat # 17117701
Superose 6 10/300 GL	GE Healthcare	Cat # 17517201
Biacore T200	GE Healthcare	Cat # 28975001
Series S sensor CM5	GE Healthcare	Cat # 29104988

Lead contact and materials availability

All resources and reagents are available from the Lead Contact, Matthew Crump (matt.crump@bristol.ac.uk).

Experimental model

Cell lines:

Spodoptera frugiperda 21 (Sf21) insect cells were grown in ESF 921 serum free medium (Expression systems) at 27 °C. Human Embryonic Kidney 293 cells were grown in Dulbecco's Modified Eagle Medium supplemented with 10 % fetal bovine serum at 37 °C in 5% CO2 incubator.

Method Details

Insect expression and purification:

The pFastBac transfer vector (Thermo Fisher Scientific) was modified to encode a 5' RPTPmu signal sequence (GILPSPGMALLSLVSLLSVLLTGVAETG) and 3' hexa-histidine tag. The genes encoding human CI-MPR domain 8 (residues 1082-1221) and domains 9-10 (residues 1222-1510, **Supplemental Table S1**) were sub-cloned into this pFastBac vector using restriction endonucleases EcoRI and HindIII (Thermo Fisher Scientific).

D8pFastBac and D9-10pFastBac was transformed into competent DH10EMBacY *E. coli* (provided by the Berger group) containing the EMBacY MultiBac bacmid. (Fitzgerald et al., 2006) Recombinant bacmid was purified by alkaline extraction and transfected (X-tremeGENE HP DNA transfection reagent, Sigma) into *Spodoptera frugiperda* 21 (Sf21) cells. First generation recombinant baculovirus was harvested 48 hrs after transfection and used to infect subsequent shaking cultures of Sf21 in ESF921 protein free media (Expression systems) at 27 °C, 120 rpm and at a density of ~ 1.0×10^6 cells/mL. Cultures were harvested 48 hours from the day of proliferation arrest - as determined by SDS-PAGE analysis and YFP emission at 524 nm.

Media containing secreted, recombinant protein was harvested by centrifugation, pH adjusted, filtered and loaded onto an Ni-NTA-agarose affinity chromatography column. The column was washed and recombinant protein eluted with an imidazole gradient. Insect expressed domain 8 and 9-10 was further purified by size exclusion chromatography (SEC) (Superdex 75 10/300 column) and eluted in 25 mM tris pH 8, 150 mM sodium chloride.

Mammalian expression and purification:

cDNA for the human CI-MPR penta-domain 7-11 (Residues 926-1649) construct was created from a larger template encoding domains 1-15 of the extracellular region of IGF2R and contained an N-terminal His$_6$ tag (**Supplemental Table S1**). DNA sequences were amplified and cloned into 293 cells by branched

PEI - 1.512 mg DNA was mixed with 67.2 mL and 6.048 mL PEI stock solution (1 mg/mL, pH 7.0) at room temperature for 8 min before adding to cell cultures. Cells were grown at 37 ºC, 5% CO_2, 5% heat-inactivated fetal bovine serum (FCS, Gibco) in 10-layer CellStack™ flasks (Corning). After 24 hours, the media was supplemented with 2% FCS, non-essential amino acids and penicillin/ streptomycin. Cultures were harvested 9 days after transfection.

For *in vivo* biotinylated versions of these proteins, the gene encoding biotin ligase was PCR amplified from the plasmid pDisplay (AddGene) and sub-cloned into pHL using restriction endonucleases MfeI and AvrII (Thermo Fisher Scientific). Genes encoding CI-MPR domains were sub-cloned into this pHL to create fusion proteins with the biotin acceptor site BirA. Constructs consisted of (5' to 3'): the *Gallus gallus* receptor-type tyrosine phosphatase S (PTPRS) signal sequence (MGFLPSPGMPALLSL VSLLSVLLMGCVA), a biotin ligase acceptor peptide, an N-terminal hexa-histidine tag and CI-MPR domains (sequences under **Supplemental Table 1**). The final construct was cloned into 293 cells and cultured as described above.

LIF was amplified by RT-PCR from human lymphocyte mRNA (a kind gift from the Neil Barclay laboratory, Sir William Dunn School of Pathology, University of Oxford). To generate mono-glycosylated LIF, glycosylation sites 2-6 were removed by site-directed mutagenesis PCR with the amino acid mutation NDQ at each site. In order to optimise signal detection in SPR experiments, LIF-N1 was sub-cloned as a fusion protein with hexa-histidine tagged mCherry using MfeI and AvrII restriction endonucleases. LIF-N1-mCherry was expressed in HEK293S cells as described above.

Media containing secreted, recombinant protein from mammalian expression was harvested and purified by SEC as described above. Mammalian expressed recombinant proteins were further purified by anion exchange chromatography followed by SEC (Superose 6 10/300 GL) into 20 mM HEPES pH 8, 200 mM NaCl.

Biophysical characterisation:
1500 µg D9-10 protein was incubated with 60 µL NP40, 60 µL NEB glycobuffer 2 and 16 µL PNGaseF (NEB) in a total volume of 600 µL for 16 hrs at 37 ºC. De-glycosylation was monitored by SDS-PAGE and mass spectrometry. De-glycosylated D9-10 was purified by analytical SEC using a Superdex 75 10/200 column (GE Healthcare).

Samples (D8 and D9-10) were prepared for mass spectrometry by methanol chloroform extraction. Data was collected on a Synapt G2-Si (Waters) electrospray ionisation–time of flight (ESI-TOF) mass spectrometer fitted with a Triverse Nanomate (Advion) spraying device in positive ion mode. Samples

were sprayed using a capillary voltage of 1.5 kV, set up for resolution mode. Data was acquired over 500-3000 m/z for 10 minutes, before analysis using MassLynx 4.1 or MagTran.

All NMR spectroscopy experiments were performed on a Bruker Avance III HD 700 MHz spectrometer equipped with a 1.7 mm inverse triple-resonance microcryocoil probe. NMR samples were prepared in 50 μL with 25 mM Tris, 150 NaCl pH 7.4 in 60 % D_2O (uncorrected for D_2O). TEAD (20 mM) was used with a final concentration of 2 mM compound. For the STD experiments (Mayer and Meyer, 2001, 1999), the standard Bruker stddiffesgp.3 pulse sequence was used with a saturation time of 7 s and a spectral width of 15.9 ppm with 256 scans. The on-resonance frequency was set to 0.58 ppm, while the off-resonance frequency was set to −28 ppm. Appropriate blank experiments, in the absence of protein or ligand, were performed to test the lack of direct saturation to the ligand protons. For the WATERLOGSY experiments (Dalvit et al., 2001, 2000), the standard Bruker ephogsygpno pulse sequence was used with relaxation delay of 1s and a mixing time of 1s with a spectral width of 15.9 ppm with 256 scans.

SPR was performed with a Biacore T200 at 25 °C. Biotinylated multi-domain constructs (D9-10, D9-11, D7-11) were immobilized at a density of 28 RU on a Series S sensor CM5, after saturation deposition of Streptavidin using NHS/EDC protocol (GE Life Sciences). The analyte, mCHERRY-LIF-N1, was added at a flow rate of 35 μl/min in HBS-EP buffer (10 mM Hepes (pH 7.4), 150 mM NaCl, 3 mM EDTA, and 0.005% (v/v) surfactant P20) for 2-3 mins, followed by a 3 min dissociation in the same buffer. Regeneration of the binding surface was achieved by injecting 2M $MgCl_2$ for 2 min, followed by a 2 min stabilization period in running buffer. Experiments were repeated twice. BIAevaluation software version 4.0.1 was used to determine kinetic parameters by global fitting of sensorgrams to the two-state (conformational change) binding model. Additionally, BIAevaluation was also used to calculate the dissociation affinity constant (K_D) by fitting the response of each concentration at equilibrium to a steady-state affinity mode

In-line SEC-SAXS of D9-10 was collected at B21 Diamond Light Source using an Agilent 1200 HPLC and 2.4 mL Superdex S200 column (GE Healthcare). 50 μL of D9-10 at 6.5 mg/mL (180 μM) was loaded onto the S200 column in running buffer (25 mM Tris, 150 mM sodium chloride) at pH 7.5. Frames were collected at 3 seconds per frame at 25 °C and X-ray scattering was recorded (Pilatus 2M detector) at a fixed camera length of 4.014 m, at 12.4 keV. Angular q range data were collected between 0.0025- 0.34 Å^{-1} Data reduction, buffer subtraction and modelling of the radius of gyration (R_g), the maximum particle dimension (D_{max}) and the pair distribution function (P(r)) were determined using ScÅtter 3.1r (Rambo, n.d.). *Ab initio* bead density shape envelope models for each dataset were generated by programs within the ATSAS 2.7,2 package (Petoukhov et al., 2012). DAMMIF (Franke, D. and Svergun, 2009) averaging over twenty three independent runs using the program DAMAVER, (Volkov and Svergun, 2003) before a single DAMMIN (Svergun, 1999) refinement run. *Ab initio* bead density shape envelope models were

superimposed to three dimensional structures of proteins using SUPCOMB. (Svergun and Kozin, 2001) FoXS and MultiFoXS (Schneidman-Duhovny D, Hammel M, Tainer JA, 2016, 2013) was used to model flexible regions and quantitatively compare the calculated X-ray scattering of three-dimensional models with the experimental scattering profile of each protein.

<u>Crystallisation, data collection and structure determination:</u>

D9-10. A sparse crystallisation screen was set up with domains 9-10 (His$_6$ tagged, 4-8 mg/mL) in Swissci 96-well plates (Molecular Dimensions) using Art Robbins Gryphon and Phoenix liquid handling robots. After ~7 weeks crystals were seen in PACT Premier condition A6 (Molecular Dimensions): 0.1 M SPG (succinic acid, sodium dihydrogen phosphate monohydrate, glycine) pH 9, 25 % PEG 1500. Suitable crystals were looped, dipped in 25 % glycerol and cryo-cooled in liquid nitrogen.

Diffraction data was collected on beamline I04 at Diamond Light Source (Didcot, UK). Diffraction images were processed and integrated in the Xia2 software package (Winter, 2010). Domain 9-10 was solved using phaser with homology models of domain 9 and domain 10 as search models. Using these phases about 90% of the model was automatically built using autobuild (Adams et al., 2010). The remaining residues were manually built in *Coot* (Emsley et al., 2010) and the model was subjected to several rounds of refinement and model building using phenix refine (Adams et al., 2010) and *Coot* (Emsley et al., 2010). Glycans were modelled using the carbohydrate module in *Coot* (Emsley et al., 2010) and validated with Privateer (Winn et al., 2011). The final model had an *R*work of 19.9% and an *R*free of 22.8% to 1.5 Å resolution. Figures were made in PyMOL ("The PyMOL molecular graphics system, version 1.5.0.4 Schrodinger, LLC," n.d.)

D8. A sparse crystallisation screen was set up with glycosylated domain 8 (His$_6$ tagged, 5-10 mg/mL) in Swissci 96-well plates (Molecular Dimensions) using Art Robbins Gryphon and Phoenix liquid handling robots. After ~6 months crystals were seen in Structure screen I+II condition E7 (Molecular Dimensions): 1.5 M Ammonium sulfate, 0.1 M Tris pH 8.5, 12 % glycerol. Suitable crystals were looped, dipped in 25 % glycerol and cryo-cooled in liquid nitrogen.

Diffraction data was collected on beamline I24 at Diamond Light Source (Didcot, UK). Diffraction images 1-958 and 1298-1800 were processed and integrated in the Xia2 software package (Winter, 2010). Domain 8 was solved using phaser with domain 10 from the D9-10 crystal structure as a search model. Using these phases about 90% of the model was automatically built using autobuild (Adams et al., 2010). The remaining residues were manually built in *Coot* (Emsley et al., 2010) and the model was subjected to several rounds of refinement and model building using phenix refine (Adams et al., 2010) and *Coot* (Emsley et al., 2010). TLS restraints were added from the TLSMD server (Painter and Merritt, 2006a,

2006b). The final model had an Rwork of 26.9 % and an Rfree of 31.0 % to 2.5 Å resolution. Figures were made in PyMOL ("The PyMOL molecular graphics system, version 1.5.0.4 Schrodinger, LLC," n.d.)

D7-11. A sparse crystallisation screen was performed with domains 7-11 (His$_6$ tagged, 3.5-7 mg/mL) in CrystalQuick 96-well plates (Greiner Bio-One). After ~20 hours crystals were observed in ProPlex condition H06 (Molecular Dimensions): 0.1 M MES pH 6.5, 1.6 M MgSO$_4$ with the addition of 10 mM M6P. These were looped, dipped in perfluoro-polyether-oil and cryo-cooled in liquid nitrogen. Many other crystal conditions were trialled with the inclusion and exclusion of mannose and M6P, but an improved resolution diffraction dataset was not achieved.

Diffraction data was collected on beamline I03 at Diamond Light Source (Didcot, UK). Diffraction images were processed and integrated by XDS (Kabsch, 2010) in the Xia2 software package (Winter, 2010). Phaser (Adams et al., 2010) was used to serially place Rosetta models of domain 7-10 and the crystal structure of domain 11 (PDB code: 1GP0). The model was built through iterative rounds of manual building in *Coot* (Emsley et al., 2010) and refinement in REFMAC5 (Vagin and Teplyakov, 2010). B-factor sharpening was used to enhance the low-resolution maps. Once the model was built, external structural restraints generated with ProSMART (Nicholls et al., 2012) using high resolution models for domains 8, 9, 10 and 11 (1.4-2.5 Å) were added to the REFMAC5 refinement (Vagin et al., 2004). Glycans were modelled using the carbohydrate module in *Coot* (Emsley et al., 2010) and validated with Privateer (Winn et al., 2011). The final model had an Rwork of 26.1% and an Rfree of 30.0% to 3.5 Å resolution. Figures were made in PyMOL ("The PyMOL molecular graphics system, version 1.5.0.4 Schrodinger, LLC," n.d.).

References:

Adams, P.D., Afonine, P. V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al., 2010. PHENIX: A comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 213–221.

Aebi, M., 2013. N-linked protein glycosylation in the ER. Biochim. Biophys. Acta 1833, 2430–2437.

Aebi, M., Bernasconi, R., Clerc, S., Molinari, M., 2010. N-glycan structures: recognition and processing in the ER. Trends Biochem. Sci. 35, 74–82.

Bieniossek, C., Richmond, T.J., Berger, I., 2008. MultiBac: Multigene baculovirus-based eukaryotic protein complex production. Curr. Protoc. Protein Sci. 5.

Bohnsack, R.N., Song, X., Olson, L.J., Kudo, M., Gotschall, R.R., Canfield, W.M., Cummings, R.D., Smith, D.F., Dahms, N.M., 2009. Cation-independent mannose 6-phosphate receptor: A composite of distinct phosphomannosyl binding sites. J. Biol. Chem. 284, 35215–35226.

Brown, J., Delaine, C., Zaccheo, O.J., Siebold, C., Gilbert, R.J., Van Boxel, G., Denley, A., Wallace, J.C., Hassan, A.B., Forbes, B.E., et al., 2008. Structure and functional analysis of the IGF-II/IGF2R interaction. EMBO J. 27, 265–276.

Brown, J., Esnouf, R.M., Jones, M.A., Linnell, J., Harlos, K., Hassan, A.B., Jones, E.Y., 2002. Structure of a functional IGF2R fragment determined from the anomalous scattering of sulfur. EMBO J. 21, 1054–1062.

Byrd, J.C., Park, J.H.Y., Schaffer, B.S., Garmroudi, F., MacDonald, R.G., 2000. Dimerization of the insulin-like growth factor II/mannose 6-phosphate receptor. J. Biol. Chem. 275, 18647–18656.

Chavez, C.A., Bohnsack, R.N., Kudo, M., Gotschall, R.R., Canfield, W.M., Dahms, N.M., 2007. Domain 5 of the cation-independent mannose 6-phosphate receptor preferentially binds phosphodiesters (mannose 6-phosphate N-acetylglucosamine ester). Biochemistry 46, 12604–12617.

Dahms, N.M., Hancock, M.K., 2002. P-type lectins. Biochim. Biophys. Acta - Gen. Subj. 1572, 317–340.

Dalvit, C., Fogliatto, G., Stewart, A., Veronesi, M., Stockman, B., 2001. WaterLOGSY as a Method for Primary NMR Screening: Practical Aspects and Range of Applicability. J. Biomol. NMR 4, 349–359.

Dalvit, C., Pevarello, P., Tato, M., Veronesi, M., Sundstrom, M., 2000. Identification of Compounds With Binding Affinity to Proteins via Magnetization Transfer From Bulk Water. J. Biomol. NMR 1, 65–68.

Devi, G.R., Byrd, J.C., Slentz, D.H., MacDonald, R.G., 1998. An insulin-like growth factor II (IGF-II) affinity-enhancing domain localized within extracytoplasmic repeat 13 of the IGF-II/mannose 6-phosphate receptor. Mol. Endocrinol. 12, 1661–1672.

Distler, J.J., Guo, J., Jourdians, G.W., 1991. The Binding Specificity of High and Low Molecular Weight Phosphomannosyl Receptors from Bovine Testes. J. Biol. Chem. 266, 21687–21692.

Dwyer, B., Lundberg, D., Iskenderian, A., Strack-logue, B., Norton, A.W., Xu, J., Meiyappan, M., Concino, M.F., Zhang, B., 2020. Expression, purification, and characterization of human mannose-6-phosphate receptor – Extra cellular domain from a stable cell line utilizing a small molecule biomimetic of the mannose-6-phosphate moiety. Protein Expr. Purif. 105589.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 486–501.

Fei, X., Connelly, C.M., MacDonald, R.G., Berkowitz, D.B., 2008. A set of phosphatase-inert "molecular rulers" to probe for bivalent mannose 6-phosphate ligand-receptor interactions. Bioorganic Med. Chem. Lett. 18, 3085–3089.

Fitzgerald, D.J., Berger, P., Schaffitzel, C., Yamada, K., Richmond, T.J., Berger, I., 2006. Protein complex expression by using multigene baculoviral vectors. Nat. Methods 3, 1021–1032.

Frago, S., Nicholls, R.D., Strickland, M., Hughes, J., Williams, C., Garner, L., Surakhy, M., Maclean, R., Rezgui, D., Prince, S.N., et al., 2016. Functional evolution of IGF2:IGF2R domain 11 binding generates novel structural interactions and a specific IGF2 antagonist. Proc. Natl. Acad. Sci. 201513023.

Franke, D. and Svergun, D.I., 2009. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. J. Appl. Crystallogr. 42, 342–346.

Franke, D., Petoukhov, M. V., Konarev, P. V., Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M., et al., 2017. ATSAS 2.8: A comprehensive

data analysis suite for small-angle scattering from macromolecular solutions. J. Appl. Crystallogr. 50, 1212–1225.

Ghobrial, G., Araujo, L., Jinwala, F., Li, S., Lee, L.Y., 2018. The Structure and Biological Function of CREG. Front. Cell Dev. Biol. 6, 1–7.

Greenall, S.A., Bentley, J.D., Pearce, L.A., Scoble, J.A., Sparrow, L.G., Bartone, N.A., Xiao, W., Baxter, R.C., Cosgrove, L.J., Adams, T.E., 2013. Biochemical characterization of individual human glycosylated pro-insulin-like growth factor (IGF)-II and big-IGF-II isoforms associated with cancer. J. Biol. Chem. 288, 59–68.

Hancock, M.K., Haskins, D.J., Sun, G., Dahms, N.M., 2002. Identification of residues essential for carbohydrate recognition by the insulin-like growth factor II/mannose 6-phosphate receptor. J. Biol. Chem. 277, 11255–11264.

Harper, J., Burns, J.L., Foulstone, E.J., Pignatelli, M., Zaina, S., Hassan, A.B., 2006. Soluble IGF2 receptor rescues ApcMin/+ intestinal adenoma progression induced by Igf2 loss of imprinting. Cancer Res. 66, 1940–1948.

Hughes, J., Surakhy, M., Can, S., Ducker, M., Davies, N., Szele, F., Bühnemann, C., Carter, E., Trikin, R., Crump, M.P., et al., 2019. Maternal transmission of an Igf2r domain 11: IGF2 binding mutant allele (Igf2r I1565A) results in partial lethality, overgrowth and intestinal adenoma progression. Sci. Rep. 9, 1–16.

Kabsch, W., 2010. XDS. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 125–132.

Kim, J.P., Olson, L.J., Dahms, N.M., 2009. Carbohydrate Recognition by the Mannose 6-phosphate Receptors. Curr. Opin. Struct. Biol. 19, 534–542.

Kreiling, J.L., Byrd, J.C., MacDonald, R.G., 2005. Domain interactions of the mannose 6-phosphate/insulin-like growth factor II receptor. J. Biol. Chem. 280, 21067–77.

Krissinel, E., Henrick, K., 2007. Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 774–797.

Leksa, V., Godár, S., Cebecauer, M., Hilgert, I., Breuss, J., Weidle, U.H., Horejsí, V., Binder, B.R., Stockinger, H., 2002. The N terminus of mannose 6-phosphate/insulin-like growth factor 2 receptor in regulation of fibrinolysis and cell migration. J. Biol. Chem. 277, 40575–40582.

Marron-Terada, P.G., Hancock, M.K., Haskins, D.J., Dahms, N.M., 2000. Recognition of Dictyostelium discoideum lysosomal enzymes is conferred by the amino-terminal carbohydrate binding site of the insulin-like growth factor II/mannose 6-phosphate receptor. Biochemistry 39, 2243–2253.

Mayer, M., Meyer, B., 1999. Characterization of ligand binding by saturation transfer difference NMR spectroscopy. Angew. Chemie - Int. Ed. 38, 1784–1788.

Mayer, M., Meyer, B., 2001. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. J. Am. Chem. Soc. 123, 6108–6117.

Moremen, K.W., Tiemeyer, M., Nairn, A. V., 2012. Vertebrate protein glycosylation: Diversity, synthesis and function. Nat. Rev. Mol. Cell Biol. 13, 448–462.

Nicholls, R.A., Long, F., Murshudov, G.N., 2012. Low-resolution refinement tools in REFMAC5. Acta Crystallogr. Sect. D Biol. Crystallogr. 68, 404–417.

Nishio, M., Umezawa, Y., Fantini, J., Weiss, M.S., Chakrabarti, P., 2014. CH-π hydrogen bonds in biological macromolecules. Phys. Chem. Chem. Phys. 16, 12648–12683.

Nykjær, A., Christensen, E.I., Vorum, H., Hager, H., Petersen, C.M., Røigaard, H., Min, H.Y., Vilhardt, F., Møller, L.B., Kornfeld, S., 1998. Receptor Targets the Urokinase Receptor to. Cell 141, 815–828.

Olson, L.J., Castonguay, A.C., Lasanajak, Y., Peterson, F.C., Cummings, R.D., Smith, D.F., Dahms, N.M., 2015a. Identification of a fourth mannose 6-phosphate binding site in the cation-independent mannose 6-phosphate receptor. Glycobiology 25, 591–606.

Olson, L.J., Dahms, N.M., Kim, J.-J.P., 2004a. The N-terminal Carbohydrate Recognition Site of the Cation-independent Mannose 6-Phosphate Receptor. J. Biol. Chem. 279, 34000–34009.

Olson, L.J., Hindsgaul, O., Dahms, N.M., Kim, J.J.P., 2008. Structural insights into the mechanism of pH-dependent ligand binding and release by the cation-dependent mannose 6-phosphate receptor. J. Biol. Chem. 283, 10124–10134.

Olson, L.J., Jensen, D.R., Volkman, B.F., Kim, J.J.P., Peterson, F.C., Gundry, R.L., Dahms, N.M., 2015b. Bacterial expression of the phosphodiester-binding site of the cation-independent mannose 6-phosphate receptor for crystallographic and NMR studies. Protein Expr. Purif. 111, 91–97.

Olson, L.J., Peterson, F.C., Castonguay, A., Bohnsack, R.N., Kudo, M., Gotschall, R.R., Canfield, W.M., Volkman, B.F., Dahms, N.M., 2010. Structural basis for recognition of phosphodiester-containing lysosomal enzymes by the cation-independent mannose 6-phosphate recepbtor. Proc. Natl. Acad. Sci. 107, 12493–12498.

Olson, L.J., Yammani, R.D., Dahms, N.M., Kim, J.-J.P.J.P., 2004b. Structure of uPAR, plasminogen, and sugar-binding sites of the 300 kDa mannose 6-phosphate receptor. EMBO J. 23, 2019–28.

Olson, L.J., Zhang, J., Dahms, N.M., Kim, J.J.P., 2002. Twists and turns of the cation-dependent mannose 6-phosphate receptor. Ligand-bound versus ligand-free receptor. J. Biol. Chem. 277, 10156–10161.

Painter, J., Merritt, E.A., 2006a. TLSMD web server for the generation of multi-group TLS models. J. Appl. Crystallogr. 109–111.

Painter, J., Merritt, E.A., 2006b. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr. - Sect. D Biol. Crystallogr. 439–450.

Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D.T., Konarev, P. V., Svergun, D.I., 2012. New developments in the ATSAS program package for small-angle scattering data analysis. J. Appl. Crystallogr. 45, 342–350.

Platt, F.M., Boland, B., van der Spoel, A.C., 2012. Lysosomal storage disorders: The cellular impact of lysosomal dysfunction. J. Cell Biol. 199, 723–734.

Probst, O.C., Puxbaum, V., Svoboda, B., Leksa, V., Stockinger, H., Mikula, M., Mikulits, W., Mach, L., 2009. The mannose 6-phosphate/insulin-like growth factor II receptor restricts the tumourigenicity and invasiveness of squamous cell carcinoma cells. Int. J. Cancer 124, 2559–2567.

Rambo, R., n.d. ScAtter. https://bl1231.als.lbl.gov/scatter/.

Sacher, M., Di Bacco, A., Lunin, V. V., Ye, Z., Wagner, J., Gill, G., Cygler, M., 2005. The crystal structure of CREG, a secreted glycoprotein involved in cellular growth and differentiation. Proc. Natl. Acad. Sci. U. S. A. 102, 18326–18331.

Saha, R.P., Bhattacharyya, R., Chakrabarti, P., 2007. Interaction Geometry Involving Planar Groups in Protein–Protein Interfaces. PROTEINS Struct. Funct. Bioinforma. 67, 84–97.

Satoh, T., Chen, Y., Hu, D., Hanashima, S., Yamamoto, K., Yamaguchi, Y., 2010. Structural Basis for Oligosaccharide Recognition of Misfolded Glycoproteins by OS-9 in ER-Associated Degradation. Mol. Cell 40, 905–916.

Satoh, T., Cowieson, N.P., Hakamata, W., Ideo, H., Fukushima, K., Kurihara, M., Kato, R., Yamashita, K., Wakatsuki, S., 2007. Structural basis for recognition of high mannose type glycoproteins by mammalian transport lectin VIP36. J. Biol. Chem. 282, 28246–28255.

Schneidman-Duhovny D, Hammel M, Tainer JA, and S.A., 2013. Accurate SAXS profile computation and its assessment by contrast variation experiments. Biophys. J. 105, 962–974.

Schneidman-Duhovny D, Hammel M, Tainer JA, and S.A., 2016. FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. Nucleic Acids Res. 44, 424–429.

Shi, X., Jarvis, D.L., 2007. Protein N-glycosylation in the baculovirus-insect cell system. Curr. Drug Targets 8, 1116–25.

Song, X., Lasanajak, Y., Olson, L.J., Boonen, M., Dahms, N.M., Kornfeld, S., Cummings, R.D., Smith, D.F., 2009. Glycan microarray analysis of P-type lectins reveals distinct phosphomannose glycan recognition. J. Biol. Chem. 284, 35201–35214.

Souza, A.T. De, Yamada, T., Mills, J.J., Jirtle, R.L., 1997. Imprinted genes in liver carcinogenesis. FASEB J. 11, 60–67.

Svergun, D.I., 1999. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys. J. 2879–2886.

Svergun, D.I., Kozin, M., 2001. Automated matching of high- and low-resolution structural models. J. Appl. Crystallogr. 34, 33–41.

Taylor, M.E., Drickamer, K., 2019. Mammalian sugar-binding receptors : known functions and unexplored roles. FEBS J. 286, 1800–1814.

The PyMOL molecular graphics system, version 1.5.0.4 Schrodinger, LLC, n.d.

Tong, P.Y., Gregory, W., Kornfeld, S., 1989. Ligand interactions of the cation-independent mannose 6-phosphate receptor. The stoichiometry of mannose 6-phosphate binding. J. Biol. Chem. 264, 7962–7969.

Vagin, A., Teplyakov, A., 2010. Molecular replacement with MOLREP. Acta Crystallogr. Sect. D Biol.

Crystallogr. 66, 22–25.

Vagin, A.A., Steiner, R.A., Lebedev, A.A., Potterton, L., McNicholas, S., Long, F., Murshudov, G.N., 2004. REFMAC5 dictionary: Organization of prior chemical knowledge and guidelines for its use. Acta Crystallogr. Sect. D Biol. Crystallogr. 60, 2184–2195.

Volkov, V., Svergun, D.I., 2003. Uniqueness of ab-initio shape determination in small-angle scattering. J. Appl. Crystallogr. 860–864.

Wang, R., Qi, X., Schmiege, P., Coutavas, E., Li, X., 2020. Marked structural rearrangement of mannose 6-phosphate/IGF2 receptor at different pH environments. Sci. Adv. 6.

Westlund, B., Dahms, N.M., Kornfeld, S., 1991. The bovine mannose 6-phosphate/insulin-like growth factor II receptor. Localization of mannose 6-phosphate binding sites to domains 1-3 and 7-11 of the extracytoplasmic region. J. Biol. Chem. 266, 23233–23239.

Williams, C., Hoppe, H., Rezgui, D., Strickland, M., Forbes, B.E., Grutzner, F., Frago, S., Ellis, R.Z., Wattana-amorn, P., Prince, S.N., et al., 2012. An Exon Splice Enhancer Primes IGF2:IGF2R Binding Site Structure and Function Evolution. Science (80-.). 338, 1209–1213.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al., 2011. Overview of the CCP4 suite and current developments. Acta Crystallogr. Sect. D Biol. Crystallogr. 67, 235–242.

Winter, G., 2010. Xia2: An expert system for macromolecular crystallography data reduction. J. Appl. Crystallogr. 43, 186–190.

Zaccheo, O.J., Prince, S.N., Miller, D.M., Williams, C., Fred Kemp, C., Brown, J., Yvonne Jones, E., Catto, L.E., Crump, M.P., Bassim Hassan, A., 2006. Kinetics of Insulin-like Growth Factor II (IGF-II) Interaction with Domain 11 of the Human IGF-II/Mannose 6-phosphate Receptor: Function of CD and AB Loop Solvent-exposed Residues. J. Mol. Biol. 359, 403–421.

Supplemental information

Structure of the human cation-independent mannose 6-phosphate/ IGF2 receptor domains 7-11 uncovers the high affinity mannose 6-phosphate binding site of domain 9.

Alice J. Bochel[1,†], Christopher Williams[1,6,†] Airlie J. McCoy[2,] , Hans-Jürgen Hoppe[3,7], Ashley J. Winter[1], Ryan D. Nicholls[1], Karl Harlos[4], E. Yvonne Jones[4], Imre Berger[5], A. Bassim Hassan[3,] and Matthew P. Crump[1,6,*]*

1. School of Chemistry, Cantock's Close, University of Bristol, Bristol BS8 1TS, UK.
2. Cambridge Institute for Medical Research, Department of Haematology, University of Cambridge, The Keith Peters Building, Hills Road, Cambridge CB2 0XY, England.
3. Tumour Growth Control Group, Oxford Molecular Pathology Institute, Sir William Dunn School of Pathology, University of Oxford, Oxford, OX1 3RE, UK.
4. Cancer Research UK Receptor Structure Research Group, Division of Structural Biology, Wellcome Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN UK.
5. School of Biochemistry, University of Bristol, University Walk, Bristol, BS8 1TD, UK.
6. BrisSynBio, Life Sciences Building, Tyndall Avenue, Bristol, BS8 1TQ, UK
7. dAInomics Ltd, c/o Diagnox, 77 Heyford Park, Upper Heyford, Oxfordshire, OX25 5HD, UK

Figure S1: The role of P-type lectins in trafficking M6P tagged ligands and internalisation of Insulin-like growth factor 2 (IGF2). A: Both P-type lectins, the CD-MPR and CI-MPR, constitutively cycle between the trans golgi network and late endosome, transporting M6P-tagged lysosomal proteins. Cargo is unloaded in the acidic environment of the late endosome and the P-type lectins recycled back to the trans golgi network. The CI-MPR is also secreted to the plasma membrane (PM) where it may bind the growth factor IGF2 independent of M6P. This results in IGF2 internalisation and subsequent lysosomal degradation (Dahms and Hancock, 2002). **B:** Generation of the M6P recognition marker occurs in two enzymatic steps. Firstly, GlcNAc-1-phosphotransferase attaches an N-acetylglucosamine (GlcNAc) 1-phosphate residue onto the terminal mannose residue of an N-linked glycan of lysosomal proteins. This creates a GlcNAc-M6P di-ester. Uncovering enzyme (UCE) removes the terminal GlcNAc exposing mannose 6-phosphate (M6P) mono-ester.(Gorelik et al., 2020; Rohrer and Kornfeld, 2001).
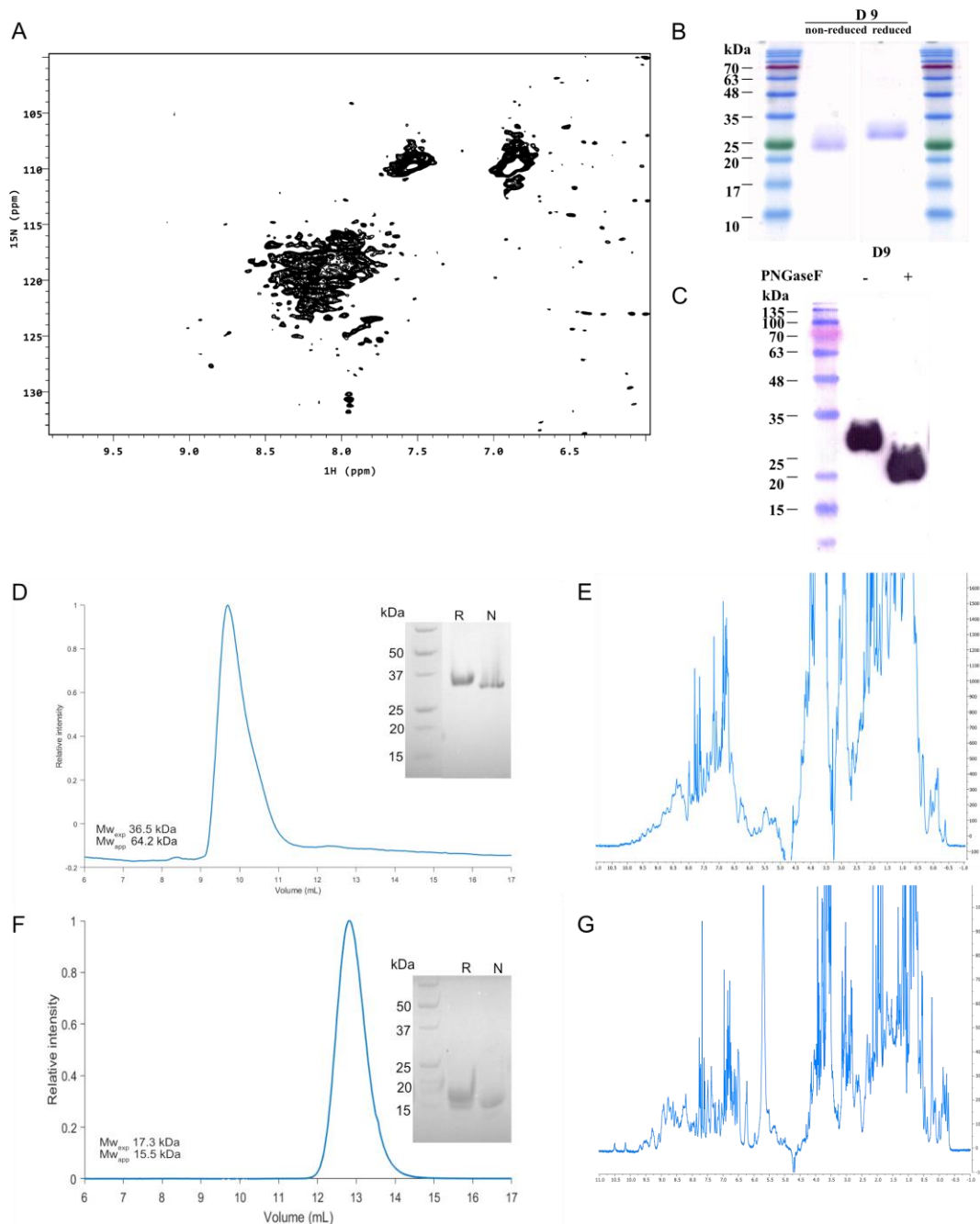
Figure S2: Initial characterisation of human CI-MPR domain 9, domains 9-10 and domain 8. A: ^{1}H-^{15}N HSQC of D9 following bacterial expression and attempted refolding. This is representative of spectra obtained following numerous refolding protocols. **B**: SDS-PAGE of purified domain 9 under reducing and non-reducing conditions. **C**: Anti-His-tag western blot of domain 9 de-glycosylated with PNGase F. **D**: Analytical gel filtration of domains 9-10 reveals dimerisation with an expected monomeric molecular weight (MW$_{exp}$) of 36.5 kDa and apparent molecular weight (Mw$_{app}$) of 64.2 kDa. SDS-PAGE analysis of peak fractions under reducing (R) and non-reducing (N) conditions. **E**: One-dimensional ^{1}H-NMR (700 MHz) of domains 9-10 confirmed protein folding. **F**: Analytical gel filtration of domain 8 reveals a monomeric status with an expected monomeric molecular weight (MW$_{exp}$) of 17.3 kDa and apparent molecular weight (Mw$_{app}$) of 15.5 kDa. SDS-PAGE analysis of peak fractions under reducing (R) and non-reducing (N) conditions. **G**: one-dimensional ^{1}H-NMR (700 MHz) of domain 8 confirmed protein folding.

Label	Species	Observed mass (Da)	Expected mass (Da)	Difference (Da)
A	Truncated D9-10 with glycans x	35813.8		
B	Truncated D9-10 with glycans 1 and 2	35975.6	35972.3	+3.3
C	Truncated D9-10 with two glycan 3	36073.6	36070.4	+3.2
D	Full length D9-10 with glycans x	36172.1		
E	Full length D9-10 with glycans 1 and 2	36333.9	36330.6	+3.3
F	Full length D9-10 with two glycan 3	36431.9	36428.7	+3.2

Label	Species	Observed mass (Da)	Expected mass (Da)	Difference (Da)
B	Truncated D9-10 with glycans 1 and 2	35972.0	35972.4	-0.4
E	Full length D9-10 with glycans 1 and 2	36331.0	36330.3	+0.7
G	Truncated D9-10 with two glycan 4	35424.0	35422.2	+1.8
H	Full length D9-10 with two glycan 4	35782.0	35780.5	+1.5

Label	Species	Observed mass (Da)	Expected mass (Da)	Difference (Da)
1	Truncated D9-10 deglycosylated	33231.0	33231.4	-0.4
2	Full length D9-10 deglycosylated	33589.0	33589.7	-0.7

| Number	Species	Observed mass (Da)	Expected mass (Da)	Difference (Da)
1	Truncated D8 minus His6 tag with 1 fucosylated glycan	16520.0	16518.5	+1.5
2	Full length D8 not glycosylated	16658.0	16660.7	-2.7
3	Full length D8 minus His6 tag with 1 fucosylated glycan	16879.0	16876.8	+1.2
4	Truncated D8 with 1 fucosylated glycan	17343.0	17341.3	+1.7
5	Full length D8 with 1 fucosylated glycan	17701.0	17699.7	+1.3

Insect glycosylation pathway

4

Figure S3: Mass spectrometry of human CI-MPR domains 9-10 and domain 8. A: ESI-MS of D9-10 used in X-ray crystallography reveals the glycosylation status of the two N-linked glycosylation sites of domain 9 (N1246 and N1312). Peak E, with mass 36333.9 Da was close to the expected mass of 36330.6 Da consistent with the presence of a low (1) and high (2) mannose glycan typical of on-pathway insect glycosylation processing (Aebi, 2013; Aebi et al., 2010; Shi and Jarvis, 2007). Similarly, the higher mass peak F (36431.9 Da) closely matched the expected mass for dual modification with hybrid glycan 3. The same glycan modifications (peaks B and C) were also observed on a second species attributed to an N-terminal truncation of D9-10 in which four amino acids (ETGA) from the N-terminal signal sequence had been lost resulting in 358.3 Da mass difference. Peaks A and D also appeared to correspond to truncated and full-length D9-10 respectively however neither mass was reconcilable with dual combinations of expected on-pathway glycans. **B**: ESI-MS analysis of D9-10 batch used in SAXS studies. **C**: ESI-MS analysis of D9-10 de-glycosylated by PNGase F. **D**: ESI-MS analysis of domain 8 reveals N-linked glycosylation at one site (N1164). The same N-terminal truncation is observed as for D9-10. **E**: The N-linked glycosylation pathway in insect cells. The pathway highlighted in red and pale green occurs in Sf21 cells used here. (Aebi, 2013; Aebi et al., 2010; Shi and Jarvis, 2007) Red boxes indicate glycans observed on domain 9.
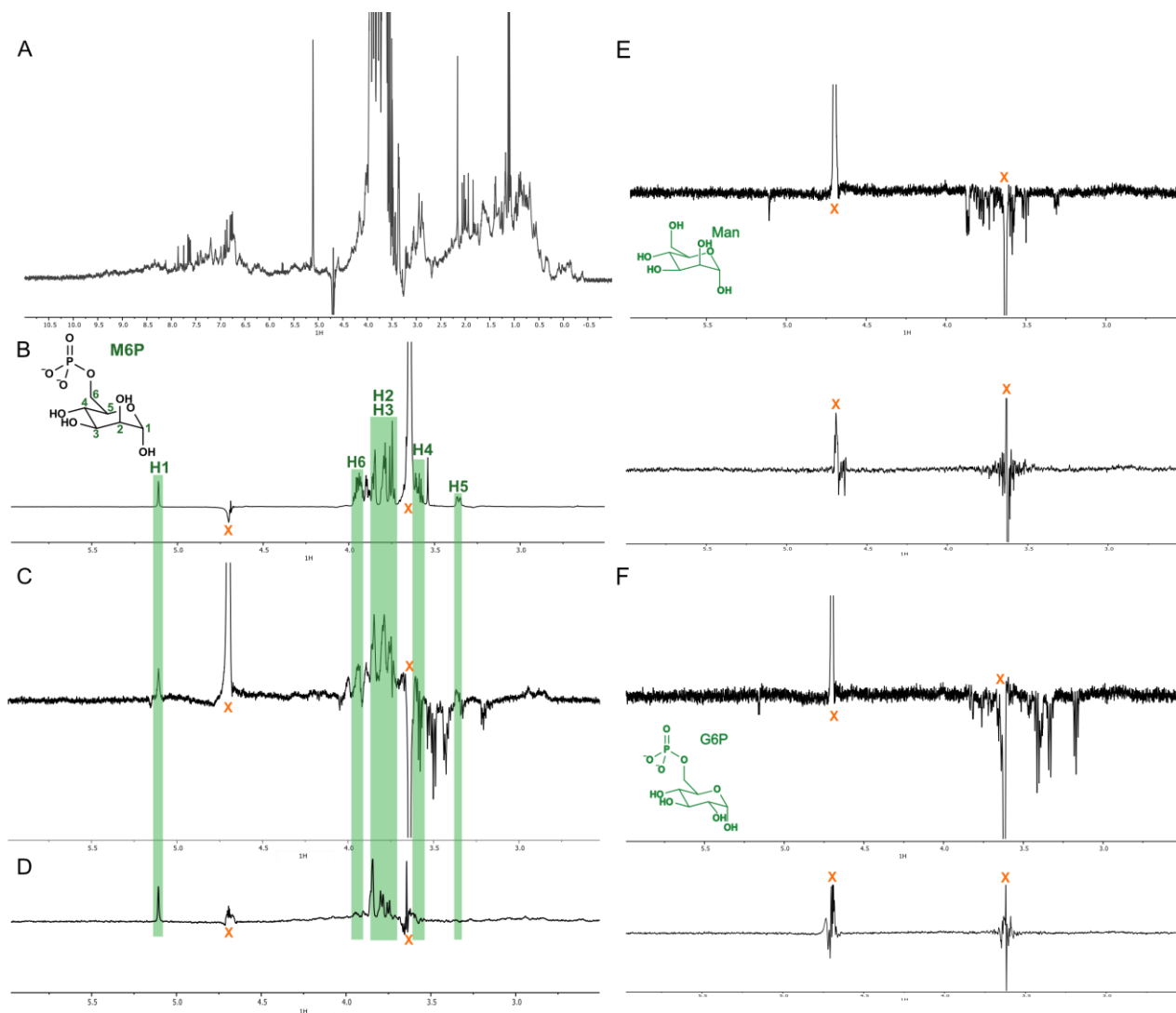
Figure S4: M6P binding to domains 9-10 construct. A: 1D ^{1}H-NMR spectrum of D9-10 plus M6P confirmed that D9-10 was folded. **B:** 1D ^{1}H-NMR spectrum of M6P. Chemical shifts of M6P protons are highlighted green. **C:** Positive signals for M6P H1, H2, H3 and H6 in the Water LOGSY of D9-10 plus M6P suggest interaction between M6P and D9-10. **D:** Positive signals in the STD difference spectrum of D9-10 plus M6P also suggest an interaction. **E:** Water LOGSY and STD spectra of D9-10 plus negative control sugar mannose, showing a lack of positive signals and thus no interaction. **F:** Similarly, the Water LOGSY and STD spectra of D9-10 plus negative control sugar glucose 6-phosphate (G6P) also demonstrates a lack of positive signals and no interaction with D9-10. All spectra were collected with 20 mM D9-10 and 2 mM M6P. STD were collected with an on-resonance frequency of 0.58 ppm and off-resonance frequency of -28 ppm Peaks marked with an orange cross correspond to water (~4.7 ppm) and tris (~3.6 ppm).
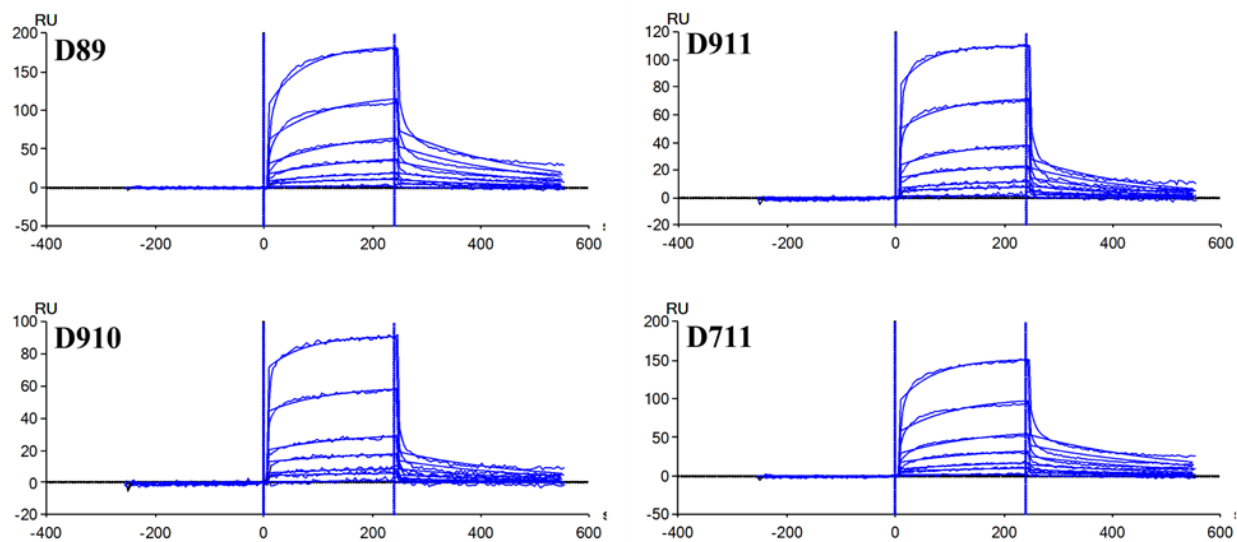
	Major binding kinetic variables (Langmuir Fitting)			
	k_a $(\times 10^5 \text{ m}^{-1}\text{ s}^{-1})$	k_d $(\times 10^{-2} \text{ s}^{-1})$	K_D $(\times 10^{-9} \text{ M})$	X^2 (RU^2)
Domain 8-9	53 ± 0.8	44 ± 4.0	81	12.1
Domain 9-10	52 ± 1.3	40 ± 0.6	77	2.2
Domain 9-11	60 ± 0.9	36 ± 3.3	60	8.5
Domain 7-11	76 ± 1.6	49 ± 0.6	65	3.7

Figure S5: M6P binding to domains 7-11. SPR analysis of domain 9 containing constructs D8-9, D9-10, D9-11 and D7-11 binding to M6P-tagged mcherry LIF-N1 give a binding affinity (K_D) of ~60-80 nM, which is in line with literature values for a single domain 9 construct binding M6P tagged GAA (K_D 95 ± 12 nM) (Chavez et al., 2007).
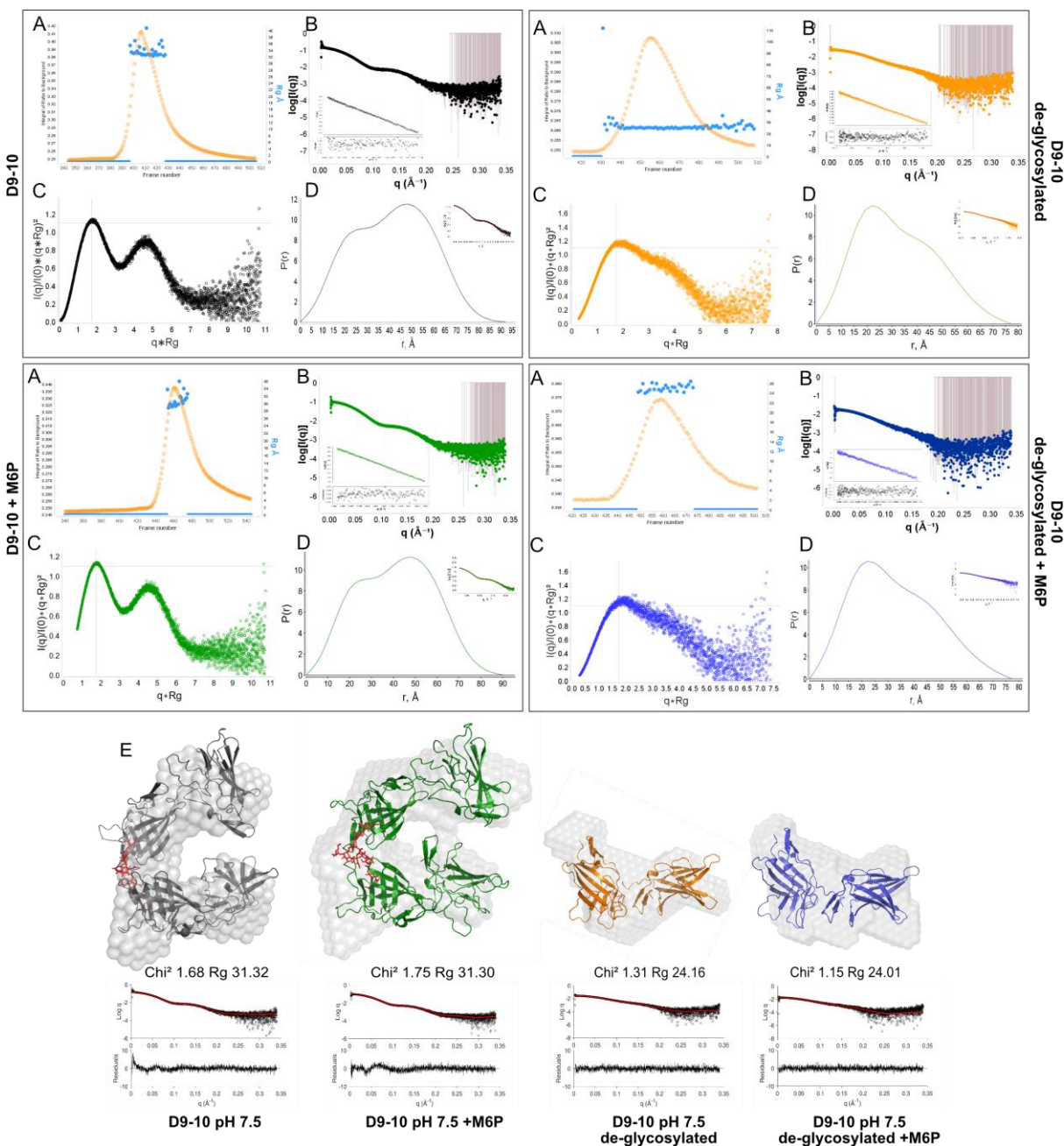
Figure S6: SEC-SAXS of D9-10 at pH 7.5 in the absence of M6P (top left, black), presence of 100-fold excess of M6P (bottom left, green), de-glycosylated in the absence of M6P (top right, orange) and presence of 100-fold excess of M6P (bottom right, blue). For each result: **A:** SEC trace of D9-10 showing the intensity of scattering as a ratio to the buffer background vs frame number. The Rg across the peak is highlighted in blue. **B**: I(q) vs q as log-linear plots with errors shown, the inset displays the Guinier fit and residuals for data at $q*R_g < 1.3$. **C:** Dimensionless Kratky plot data shown in panel **B**. The crosshair (1.10, 1.73) is the expected peak maxima for a folded globular protein. The two peaks observed indicate that the species is a two-domain structure. The definition of the second peak is reduced when the sample is deglycosylated. **D:** Pair-distance, P(r) distribution function showing the P(r) vs r profile. The inset shows the fit of the P(r) distribution profile (red line) to the raw scattering data used in the P(r) vs r analyses. **E:** *Ab initio* DAMMIN bead density shape envelopes of D9-10 are represented as a surface (light grey) overlaid with a 1-state-optimised model. As D9-10 from the D7-11 crystal structure was used as a starting model there is no His tag or second glycan (except for D9-10 pH 5.5 where the second glycan at N1246 was built in and for the de-glycosylated samples where the glycan was deleted). Applicable glycans are shown in red.

8
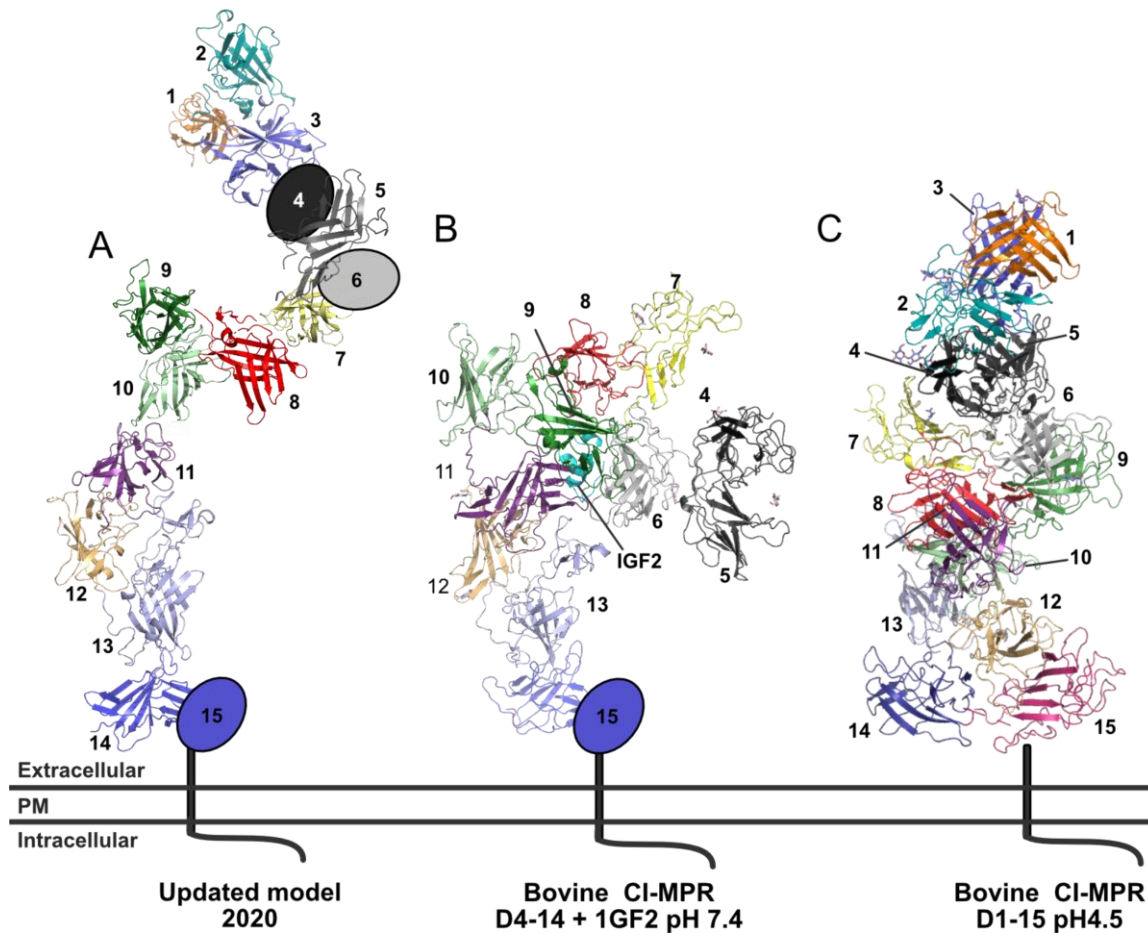
Figure S7: The structure of the extracellular region of the CI-MPR. A: Chain A of our proposed model of the CI-MPR D1-14 based upon the structures of D1-3, D5, D7-11, D11-14. There are no high-resolution structures of D4, D6 and D15. **B:** Cryo-EM structure of Bovine CI-MPR D1-14 at pH 4.5 (3.46 Å). **C:** Cryo-EM structure of Bovine CI-MPR D4-14 in complex with IGF2 at pH 7.4 (4.32 Å). PDB codes: 6UM1 and 6UM2 respectively (Wang et al., 2020).

D8			
	ETGASVDCQVTDLAGNEYDLTGLSTVRKPWTAVDTSVDGRKRTFYLSVCNPL	1227	D8
1128	PYIPGCQGSAVGSCLVSEGNSWNLGVVQMSPQAAANGSLSIMYVNGDKCGNQRFSTRITF	1188	D8
1189	ECAQISGSPAFQLQDGCEYVFIWRTVEACPVVRTRHHHHHH*		D8
D9-10			
	ETGASVEGDNCEVKDPRHGNLYDLKPLGLNDT	1248	D9
1249	IVSAGEYTYYFRVCGKLSSDVCPTSDKSKVVSSCQEKREPQGFHKVAGLLTQKLTYENGL	1308	D9
1309	LKMNFTGGDTCHKVYQRSTAIFFYCDRGTQRPVFLKETSDCSYLFEWRTQYACPPFDLTE	1368	D9/10
1369	CSFKDGAGNSFDLSSLSRYSDNWEAITGTGDPEHYLINVCKSLAPQAGTEPCPPEAAACL	1428	D10
1429	LGGSKPVNLGRVRDGPQWRDGIIVLKYVDGDLCPDGIRKKSTTIRFTCSESQVNSRPMFI	1488	D10
1489	SAVEDCEYTFAWPTATACPMKSTRHHHHHH*		D10
D7-11			
	QLKHHHHHHEFTTDTDQACSIRDPNSGFVFNLNP	949	D7
950	LNSSQGYNVSGIGKIFMFNVCGTMPVCGTILGKPASGCEAETQTEELKNWKPARPVGIEK	1009	D7
1010	SLQLSTEGFITLTYKGPLSAKGTADAFIVRFVCNDDVYSGPLKFLHQDIDSGQGIRNTYF	1069	D7
1070	EFETALACVPSPVDCQVTDLAGNEYDLTGLSTVRKPWTAVDTSVDGRKRTFYLSVCNPLP	1129	D7/8
1130	YIPGCQGSAVGSCLVSEGNSWNLGVVQMSPQAAANGSLSIMYVNGDKCGNQRFSTRITFE	1189	D8
1190	CAQISGSPAFQLQDGCEYVFIWRTVEACPVVRVEGDNCEVKDPRHGNLYDLKPLGLNDTI	1249	D8/9
1250	VSAGEYTYYFRVCGKLSSDVCPTSDKSKVVSSCQEKREPQGFHKVAGLLTQKLTYENGLL	1309	D9
1310	KMNFTGGDTCHKVYQRSTAIFFYCDRGTQRPVFLKETSDCSYLFEWRTQYACPPFDLTEC	1369	D9/10
1370	SFKDGAGNSFDLSSLSRYSDNWEAITGTGDPEHYLINVCKSLAPQAGTEPCPPEAAACLL	1429	D10
1430	GGSKPVNLGRVRDGPQWRDGIIVLKYVDGDLCPDGIRKKSTTIRFTCSESQVNSRPMFIS	1489	D10
1490	AVEDCEYTFAWPTATACPMKSNEHDDCQVTNPSTGHLFDLSSLSGRAGFTAAYSEKGLVY	1549	D10/11
1550	MSICGENENCPPGVGACFGQTRISVGKANKRLRYVDQVLQLVYKDGSPCPSKSGLSYKSV	1609	D11
1610	VISFVCRPEARPTNRPMLISLDKQTCTLFFSWHTPLACEQA*		D11
BirA			
	ETGQLKDNTVPLKLIALLANGEFHSGEQLGETLGMSRAAINKHIQTLRDWGVDVFTVPGK		
	GYSLPEPIQLLNAKQILGQLDGGSVAVLPVIDSTNQYLLDRIGELKSGDACIAEYQQAGR		
	GRRGRKWFSPFGANLYLSMFWRLEQGPAAAIGLSLVIGIVMAEVLRKLGADKVRVRWPND		
	LYLQDRKLAGILVELTGKTGDAAQIVIGAGINMAMRRVEESVVNQGWITLQEAGINLDRN		
	TLAAMLIRELRAALELFEQEGLAPYLSRWEKLDNFINRPVKLIIGDKEIFGISRGIDKQC		
	ALLLEQDGIIKPWMGGEISLRSAEKGSTSGSGGSKDEL*		

Table S1: Amino acid sequences of human CI-MPR D8, D9-10, D7-11 and BirA. D8 and D9-10, which were expressed in Sf21 insect cells, contain an N-terminal signal sequence (residual sequence in black) and a C-terminal His-tag (red). D7-11, which was expressed in HEK 293S mammalian cells, contains an N-terminal His-tag. CI-MPR domain 7 (T9284-P1081) is coloured orange, domain 8 (V1082-R1221) green, D9 (V1222-D1365) navy blue, D10 (L1366-S1510) blue and D11 (N1511-T1651) grey. D7 contains three disulphide bonds (underlined) while D8, 9, 10 and 11 each contain four disulphide bonds. BirA (navy blue) contains an N- and C-terminal linker (black) and was inserted at the N-terminal of the boundaries shown for the D7-11 construct and included BirA-D8-9, BirA-D9-10, BirA-D9-11 and BirA-D7-11 (replacing the His-tag)

Construct	Domains 9-10	Domain 8	Domains 7-11
Data collection			
PDB accession code	**6Z30**	**6Z31**	**6Z32**
Space group	P 2_1 2_1 2_1	P$12_1$1	P 4_1 2_1 2
Unit cell			
a, b, c (Å)	40.4, 55.8, 132.7	65.6, 45.6, 70.3	139.2 139.2 234.7
α, β, γ (°)	90, 90, 90	90, 116, 90	90, 90, 90
X-ray wavelength (Å)	0.98	0.97	1.06
Resolution range (Å)	51.46-1.50 (1.55-1.50)	36.02-2.56 (2.65-2.56)	89.72-3.47 (3.59-3.47)
Total reflections	630776 (64,630)	78211 (7481)	438,999 (43,737)
Unique reflections	48,880 (4805)	12129 (1183)	30,224 (2944)
Multiplicity	12.9 (13.5)	6.4 (6.4)	14.5 (14.9)
Completeness (%)	99.9 (99.9)	97.9 (93.4)	98.6 (98.4)
R_{meas}	0.15 (2.4)	0.62 (1.97)	0.21 (5.06)
Mean I/σ (I)	11.12 (1.24)	5.02 (2.56)	6.03 (0.84)
Wilson B-factor (Å^2)	21.15	16.61	157.79
CC ½	0.999 (0.488)	0.973 (0.753)	0.997 (0.127)
Refinement			
Reflections used in refinement	48877 (4805)	12000 (1137)	30215 (2943)
R_{work} (%)	19.9 (30.02)	26.9 (32.0)	26.1 (42.1)
R_{free} (%)	22.8 (32.0)	31.0 (36.8)	30.0 (43.2)
Root mean squared deviation			
Bond lengths (Å)	0.005	0.010	0.014
Bond angles (°)	0.84	1.12	1.97
Ramachandran plot (%)			
Favoured	98.6	94.9	94.5
Allowed	1.4	4.4	5.1
Outliers	0.0	0.7	0.4
Average B-factor (Å^2)	27.43	21.75	197.43
Protein	25.78	21.74	197.85
Ligand/ glycan	22.68	37.76	174.54

Table S2: Data collection and structure refinement statistics for human CI-MPR domains 9-10, domain 8 and domains 7-11. Values in parentheses are for the outer resolution shell.

Data collection parameters				
Sample	**D9-10 pH 7.5**	**D9-10 pH 7.5 + M6P**	**D9-10 pH 7.5 de-glycosylated**	**D9-10 pH 7.5 de-glycosylated +M6P**
SASBDB code	**SASDH79**	**SASDH59**	**SASDH69**	**SASDJ23**
Instrument	SEC-SAXS at B21 Diamond Light Source			
SEC Column	Superdex 200 Increase 3.2/300			
Loading concentration (mg ml^{-1})	6.5			
Injection volume (μl)	45			
Flow rate (ml min^{-1})	0.075			
Average C in combined data frames (mg ml^{-1})	1.65			
Solvent	25 mM Tris-HCl, 150 mM NaCl			

Data analysis				
Guinier analysis:				
$I(0)$ (cm^{-1})	1.37E-1 ±1.4E-4	9.72E-2 ±1.3E-4	2.74E-2 ±2.3E-4	1.74E-2 ±2.9E-5
Rg (Å)	32.79±0.19	32.51±0.19	25.21±0.13	25.05±0.21
q_{min} (Å^{-1})	0.015	0.022	0.011	0.012
qR_g max (q_{min} max)	1.25 (0.034)	1.27 (0.034)	1.29 (0.040)	1.29 (0.041)
Coefficient of Correlation	0.9980	0.9986	0.9961	0.9898
P(r) analysis:				
$I(0)$ (cm^{-1})	1.31E-1 ±1.5E-3	9.29E-2 ±1.3E-3	2.76E-2 ±1.8E-4	1.79E-2 ±1.6E-4
Rg (Å)	31.38±0.20	31.16±0.18	25.23±0.11	25.32±0.13
d_{max} (Å)	92.0	91.0	77.0	77.5
q range (Å^{-1})	0.0026-0.34	0.0026-0.34	0.0026-0.34	0.0026-0.34
χ^2	0.77	0.92	1.18	1.06
Porod volume (Å^{-3})	135,067	127,660	68,881	69514
(ratio V_P/calculated M)	(2.01)	(1.87)	(2.05)	(2.07)
Resolution (from SASRES) (Å)	39 ± 3	37 ± 3	30 ±2	34 ±3
DATCLASS designation	Flat	Flat	Flat	Compact

Molecular mass determination (ratio to predicted value)				
Partial specific volume ($\bar{v}$, cm^{3}g^{-1})	0.726			
Particle contrast from sequence and solvent constituents $\Delta\bar{p}$ ($P_{protein} - P_{solvent}$; 10^{10} cm^{-2})	3.049 (12.462 – 9.413)			
Mr from sequence (KDa)	33.6	33.6	33.6	33.6
Mr form I(0) Reciprocal (KDa)	83.6 (1.24)	78.9 (1.17)	48.0 (1.43)	41.8 (1.24)
Mr from I(0) Real (KDa)	64.0 (0.95)	62.0 (0.92)	47.7 (1.42)	43.3 (1.28)
Mr from Porod volume (Vporod/1.7) (KDa)	79.5 (1.18)	75.1 (1.12)	40.7 (1.21)	42.3 (1.26)
Mr from SAXSMoW (KDa)	70.5 (1.05)	68.0 (1.01)	33.7 (1.01)	34.8 (1.04)

Shape model fitting results				
DAMMIF (default parameters, 23 calculations)				
q range for fitting (Å^{-1})	0.0026-0.233	0.0026-0.233	0.0026-0.233	0.0026-0.239
Symmetry, anisotropy assumptions	P1, none	P1, none	P1, none	P1, none
NSD (standard deviation)	0.766 (0.021)	0.612 (0.039)	0.548 (0.019)	0.597 (0.018)
DAMMIN (default parameters, slow run)				
q range for fitting (Å^{-1})	0.0026-0.233	0.0026-0.233	0.0026-0.233	0.0026-0.233
Symmetry, anisotropy assumptions	P1, none	P1, none	P1, none	P1, none
χ^2	1.937	2.158	1.504	1.320

Atomistic Modelling				
Model	**6Z32**	**6Z32**	**6Z30**	**6Z30**
q range for all modelling	0.0026-0.34	0.0026-0.34	0.0026-0.34	0.0026-0.34
FoXS:				
χ^2	3.56	9.90	2.13	1.35
Predicted R$_g$ (Å)	31.21	31.21	23.06	23.06

MultiFoXS multistate models: (10,000 iterations)				
Flexible residues	1358-1364	1358-1364	1358-1364	1358-1364
No of states	1	1	1	1
χ^2	1.68	1.75	1.31	1.15
c_1, c_2	1.03, 0.72	1.02, 1.53	0.99, 2.92	0.99, 2.74
R_g values of each state (Å)	31.32	31.30	24.15	24.01

Table S3: Data collection parameters and statistics obtained from SEC-SAXS experiments of D9-10 at pH7.5 in the absence of M6P, presence of 100-fold excess M6P, de-glycosylated in the absence of M6P and presence of M6P. For glycosylated samples D9-10 from the D7-11 crystal structure, which was singly glycosylated, was used as a starting model. All glycan residues were also specified as flexible during multistate modelling. For de-glycosylated samples D9-10 from the D9-10 crystal structure, with glycans removed, was used as the starting model.

Supplemental Experimental Procedures.

CI-MPR multi-species Domain 9 expression and purification in *E. coli*.

CI-MPR domain 9 (residues 1221-1365) was amplified from human CI-MPR cDNA and cloned into pET26b (Novagen) using *Nde*I and *Xho*I. 3' primers with or without a stop codon were used to generate non-tagged and C-terminal His$_6$-tagged constructs.

Protein was expressed in *E. coli* BL21(DE3) in LB or 2YT with 0.5 mM IPTG overnight at 25 ℃. As for other IGF2R domains purified to date the protein formed inclusion bodies which were purified using a standard CI-MPR domain 11 protocol. (Williams et al., 2012) Briefly, following induction the cells were harvested and resuspended in lysis buffer (50 mM Tris pH 8.0, 0.5 M NaCl, 10% glycerol and 0.1% Triton X-100) and lysed by sonication. The suspension was then centrifuged at 9000 g for 20 minutes to separate the soluble material from the inclusion bodies. Inclusion bodies were further purified by repeated washes in 50 mM Tris pH 8.0, 0.5 M NaCl, 1% Triton X-100, 2 M Urea, 1 mM EDTA and 5 mM β-mercaptoethanol and 50 mM Tris pH 8.0, 0.1 M NaCl, 2 M Urea, 1 mM EDTA and 5 mM β-mercaptoethanol. Pure inclusion bodies were then dissolved in denaturation buffer (0.1 M Tris pH 8.0, 8 M Urea, 0.1 M NaCl, 10 mM DTT, and 10 mM EDTA).

Initially both IGF2R domain 9 constructs were refolded following the CI-MPR domain 11 protocol. (Williams et al., 2012) Denatured protein was diluted 50:50 with fresh denaturation buffer and reduced with 10 mM fresh DTT on ice for 1 hour. The protein solution was then spun down at 13000 g for 20 minutes to remove insoluble material and added drop wise to chilled refold buffer (0.1 M Tris pH 8.5, 1 M L-arginine, 1 mM EDTA, 6.5 mM cysteamine, 3.7 mM cystamine) with vigorous stirring. The refold was then incubated at 4 ℃ with gentle stirring. After 24 or 48 hours the protein was concentrated in a stirred cell with a 10 kDa MWCO membrane and purified by gel filtration into 20 mM Tris pH 8.0, 0.5 M NaCl. Typically, 10-20 mg of denatured protein was refolded at a final concentration of 0.1 to 0.01 mg/ml.

Refolding CI-MPR domain 9 generally yielded soluble protein, however this eluted at a volume larger than expected for a well-folded protein on a gel filtration column. Further analysis using ^{1}H-^{15}N HSQCs on ^{15}N labelled material showed this fraction was misfolded. To obtain correctly folded protein the refold conditions were extensively optimised by varying the pH, buffer and final protein concentration. The concentration of arginine was also screened as were other folding additives such as glycerol, urea, detergents (sarkosyl/CHAPS), alcohols and the redox pair/ratio. Refolding purified protein from denatured Ni^{2+} affinity chromatography, either by dilution or on-column also failed to produce correctly folded protein.

Despite extensive optimisation, correctly folded protein was not obtained which may have been due to incorrect disulphide bond shuffling during the refold process. Previously, the *E. coli* disulphide bond proteins (Dsb) have been successfully used to express soluble disulphide rich proteins in *E. coli*. (Kurokawa et al., 2000) To this end it was thought co-expressing CI-MPR domain 9 with *E. coli* Dsb ABCD may aid correct disulphide bond formation during expression. The CI-MPR domain 9 gene was subsequently recloned into pET26b using *Nco*I and *Xho*I to generate CI-MPR domain 9 N-terminally fused to a pelB leader sequence for periplasmic localisation. This vector was transformed into *E. coli* BL21(DE3) along with pDsbABCD. (Kurokawa et al., 2000) Double resistant colonies were cultured in LB medium at 37 ℃, 200 RPM until exponential growth phase was reached, at which point DsbABCD expression was induced with 1 mM L-arabinose. One hour later D9 expression was induced with 0.5 mM IPTG and the culture was incubated at 25 ℃, 200 RPM overnight before harvest. The periplasmic fraction was extracted using osmotic shock as described previously (Jonathan T. Sockolosky and Francis C. Szoka, 2013) and purified by Ni^{2+} affinity chromatography. The insoluble material was purified as

described above. Unfortunately, CI-MPR domain 9 was found exclusively in the insoluble fraction, presumably nascent polypeptide aggregates immediately upon translation unless protected by glycosylation or chaperones.

CI-MPR is relatively well conserved in mammals and domain 9 sequences from other species may be more amenable for expression and refolding using an *E. coli* expression system. Therefore, as an alternative strategy CI-MPR domain 9 from *Bos taurus* (residues 1232 to 1372), *Lemur catta* (residues 1153 to 1274) and *Ornithorhynchus anatinus* (duckbill platypus; residues 1214 to 1355) were synthesised (GeneArt) and cloned into pET28a using *Nde*I and *Xho*I. The proteins were expressed in *E. coli* BL21(DE3) and refolded as described above. However, as for the human construct these constructs also proved intractable to refolding.